Optimization of the Valence Energy Variance of the CuH molecule.

Peter Belohorec, B.Sc. Physics

A thesis
submitted to the Department of Physics
in partial fulfilment of the requirements of the
degree of Master of Science.

November 1992
Brock University
St.Catharines, Ontario

To my wife
and to my parents.

Acknowledgements.

I gratefully acknowledge generous support as well as helpful discussions and instructions from Professors S.M.Rothstein and J.Vrbik during my research on this project. I would also like to thank them for their careful reading of my thesis and for their valuable suggestions as well as Peter Harris for his proofreading of this work.

I am pleased to acknowledge receipt of The Thompson-Harrison Graduate Scholarship, which was of great assistance to me in my studies.

Last but not least I would like to thank to my wife Katarina for her enormous support which accompanied me throughout my studies.

Contents:

Abstract

We developed the concept of split-τ to deal with the large molecules (in terms of the number of electrons and nuclear charge Z). This naturally leads to partitioning the local energy into components due to each electron shell. The minimization of the variation of the valence shell local energy is used to optimize a simple two parameter CuH wave function.

Molecular properties (spectroscopic constants and the dipole moment) are calculated for the optimized and nearly optimized wave functions using the Variational Quantum Monte Carlo method. Our best results are comparable to those from the single and double configuration interaction (SDCI) method.

1.1. Introduction to Monte Carlo.

Monte Carlo (MC) methods, in general, are those which are based on using random numbers. These methods are capable of solving different kinds of problems in physics, chemistry and in other areas of science. Using MC methods one can study phenomena such as an interaction of atomic and subatomic particles or even the large scale behavior of the social system, by simulating them using random numbers. The ideas that are behind MC methods already existed in the 1900s but scientists were able to exploit them only when the first fast computing machines became available.

Random numbers in MC methods were first used for solving problems arising from **probabilistic** contexts. Problems of this kind often involve simulations of a large number of particles, accounting for the forces between the particles and then computing the averages of different properties of the system. This approach, say in the theories of liquids, assumes that classical statistics are involved as well as two-body interactions and that the potential field of the molecule is spherically symmetric. Other assumptions may be made according to the nature of the problem.

In 1949 a Symposium on Monte Carlo Methods [1] was held and a distinction was made between the above mentioned kinds of MC methods and a new one which was defined as an application of stochastic methods to problems arising from the **nonprobabilistic** context. What is basically meant by this is that the new problems involved the reformulation of the equations to the form which can be treated using the diffusion of particles of different types in the configuration space of the system.

The law of motion which is the most important for physicists who deal with atomic particles is the time-dependent Schrodinger equation (SchE). In atomic units ($h/2\pi=1, m_e=1$) it has the form

$$i\frac{\partial \Psi}{\partial t} = H\Psi \tag{1}$$

This differential equation in general contains all the interactions of the system and specifies how the quantum system evolves. What is, however, usually an object of interest is the time-independent SchE (eq.(2)), which tells us what energy values are allowed for the system and what are the corresponding eigenfunctions, since these quantities are fundamental for the description of the system.

$$H\Psi = E\Psi \tag{2}$$

This equation is, for all systems of interest, a very complicated second order linear partial differential equation which can be solved only approximately. There are many methods and approximations used, but until 1949 these all arised from the non-probabilistic form of this equation. The first one who noted the similarity of the time-dependent SchE to a diffusion equation, and proposed that a suitable random walk could be performed to solve this equation was Fermi [1]. Metropolis and Ulam followed his proposal, formulated a random walk on a lattice for a harmonic oscillator and one-dimensional coulomb-like potential and used the MC method to compute the ground state energy [2,1] for both systems. That was the beginning of Quantum Monte Carlo (QMC).

There are in principle three different types of QMC methods. One of them, called **Variational Quantum Monte Carlo** (VQMC), uses a random walk procedure to evaluate numerically the expectation values for different operators given by integrals, applied to a given trial wave function Ψ_T which is known before the computation. The other type of method **solves** SchE exactly without the requirement to have highly accurate trial wave functions prior to the computation in order to determine energy or other molecular properties. These methods were, however, not capable of providing competitive results to the standard techniques until a new approach, suggested by Kalos came on the scene in the 1960's [3,4]. The proposed sophisticated method is known as **Green's Function Monte**

Carlo (GFMC). Yet another method was developed in the 1970's by Anderson [5] which is called **Diffusion Quantum Monte Carlo** (DQMC) and which solves problems in quantum chemistry using random walks as well. This method was later improved by Anderson, Moskowitz, Kalos, Ceperley, Reynolds and Lester Jr. [6-8]. Both DQMC and GFMC methods use Green's functions for solving the time-independent SchE, therefore we could use the abbreviation GFMC for both of them. However, this term usually denotes the method which uses the **exact** propagator. We will describe some differences between these two approaches later. The short introduction of all the basic QMC methods which are in use nowadays would not be complete without mentioning **Path Integral Quantum Monte Carlo** (PIQMC). It differs from all the previously mentioned methods in the fundamental way in which the kinetic energy operator is treated. In both DQMC and GFMC it is done by a random diffusion of walkers through the configuration space, whereas in PIQMC the kinetic energy operator becomes part of the total "potential" energy.

Calculations of different molecular properties such as energies, dipole moments, binding energies and charge distributions etc., are needed in quantum chemistry. Different methods are in use nowadays to achieve these goals and QMC is one of them. It has an advantage over other routinely used methods like Hartree-Fock (HF), Multiconfiguration Self-Consistent Field (MCSCF), Configuration Interaction (CI) (not Full CI) or some other Many-Body Perturbation Techniques (MBPT) in that QMC in principle can give exact results. Now we will mention some drawbacks of CI and QMC to compare them.

In CI methods, wave functions routinely account for 75% of the correlation energy [9], but this is not always sufficient because chemical effects occur on an energy scale which is much smaller than this. The other thing which must be mentioned is the CPU time dependence on the number of electrons (N) of the molecule, which is usually in the range N^4 - N^5 [10]. These CPU time costs are usually cited for fixed approximations thus compare only the amount of CPU time needed to perform all the calculations involved in a certain method without particular emphasis on the accuracy. For the evaluation of CPU time costs **at fixed accuracy** we must have in mind, that the complexity of the wave function and therefore the processor time grows like N! [10].

DQMC or GFMC on the other hand can in principle yield the exact stochastic solution to the SchE and thus the only limitations on the accuracy are due to the statistical errors involved in the method. Properties of interest are computed during the simulation as soon as the stationary state is obtained (no large scale changes are visible, only statistical fluctuations). Averages of properties of interest will give us the expectation values. One of the aims in QMC calculations is reduction of statistical errors which can be done by importance sampling using a guiding function which should be as close to the exact wavefunction as possible. CPU time in case of QMC methods increases as N^3 only, which is much better than in the above mentioned CI method. This gives us a promise of effective and fast calculation of different properties for large molecules. There are, however, limitations of QMC as well. These are especially:

1) large CPU requirements for making the statistical errors sufficiently small to get exact results or the results which show at least all the interesting effects on the chemistry energy scale;

2) lack of stable algorithms for fermions which have to be treated using fixed node approximations because of their statistics;

3) difficulty of simulating atoms with large atomic numbers because of a **critical slowing down** effect which is caused by steep dependence of CPU time on the nuclear charge Z [11,13], to reduce the statistical uncertainties to the level of chemical accuracy.

1.2. Diffusion Quantum Monte Carlo.

Now we would like to describe two QMC methods, namely Diffusion Quantum Monte Carlo and Green's function Monte Carlo, their similarities and their differences as well.

We have mentioned previously that the basic equation in quantum chemistry (the law of motion for atomic particles) is the non-relativistic fixed-nuclei Schrodinger equation, which in atomic units has the following form.

$$i \frac{\partial \Psi}{\partial t} = (- \frac{1}{2} \Delta + V(\vec{r})) \Psi \qquad (1)$$

Here $\Psi(\vec{r},t)$ is the wavefunction of 3N+1 variables, since the vector $\vec{r}$ represents all the spatial coordinates of N electrons and the potential energy operator is given by

$$V(\vec{r}) = \sum_{i<j} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{i\alpha} \frac{Z\alpha}{|\vec{r}_i - \mathbf{R}\alpha|} + \sum_{\alpha<\beta} \frac{Z\alpha Z\beta}{|\mathbf{R}\beta - \mathbf{R}\alpha|} \qquad (2)$$

where $\vec{r}_i$ means the position of the i-th electron and $\mathbf{R}\alpha$ is the fixed position of the α-th nucleus. The symbol Δ means the Laplacian of all the electronic coordinates.

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \ldots + \frac{\partial^2}{\partial x_{3N}^2} \qquad (3)$$

The formal solution for this time-dependent Schrodinger equation (1)

is given by the following infinite sum

$$\sum_{n=0}^{\infty} a_n \, \phi_n(\vec{r}) \, e^{-itE_n} \tag{4}$$

where the coefficients $\{a_n\}$ can be found from the initial condition at time t=0 and the functions $\phi_n(\vec{r})$ are the eigenfunctions of the Hamiltonian corresponding to the eigenvalues E_n.

$$\left(-\frac{1}{2}\Delta + V(\vec{r}) \right) \phi_n(\vec{r}) = E_n \, \phi_n(\vec{r}) \tag{5}$$

However, what we usually want to find is just the part of the set $\{\phi_n(\vec{r}), E_n\}$ corresponding to the ground state, namely $\{\phi_0(r), E_0\}$. The solution of another type of time dependent equation namely

$$\frac{\partial \Psi}{\partial \tau} = \left(\frac{1}{2}\Delta - (V(\vec{r}) - E_T) \right) \Psi \tag{6}$$

is given by the following infinite sum

$$\Psi(\vec{r},\tau) = \sum_{n=0}^{\infty} a_n \, \phi_n(\vec{r}) \, e^{-\tau(E_n - E_T)} \tag{7}$$

where $\phi_n(\vec{r})$ and E_n are the solutions of the same time-independent equation (5) . The only difference is in the energy shift in the Hamiltonian given by E_T, where E_T is an arbitrary number. We can easily use equation (7) for finding the lowest energy of the system E_0 (and the corresponding eigenstate $\phi_0(\vec{r})$) by using a procedure which takes a limit $\tau \to \infty$, and simultaneously adjusts $E_T \to E_0$, so that we finally have

$$\Psi(\vec{r},\tau\to\infty) = a_0\phi_0(\vec{r})+a_1\phi_1(\vec{r})e^{-\tau(E_1-E_0)}+... \qquad (8)$$

If there is sufficiently large gap in energy between the ground state and the first excited state, and the initial wave function $\Psi(\vec{r},t=0)$ has sufficiently large overlap with the ground state ($a_0 \neq 0$), then in the limit $\tau\to\infty$ the only term which survives, namely $a_0\phi_0(\vec{r})$, is proportional to the ground state.

Equation (6) has a formal operator solution of the form

$$\hat{G} = e^{-\tau\hat{H}} \qquad (9)$$

which gives us the functional solution at any time τ if we know the solution at zero time (τ=0)

$$\Psi(\tau) = \hat{G}\Psi(0) \qquad (10)$$

We can rewrite it for better understanding in the $\vec{r}$-representation in the following way

$$\langle\vec{r}\,|\Psi(\tau)\rangle = \int\langle\vec{r}|e^{-\tau\hat{H}}|\vec{r}'\rangle\langle\vec{r}'|\Psi(0)\rangle\, d^3\vec{r}'$$

which is

$$\Psi(\vec{r}\,',\tau) = \int G(\vec{r}\;,\vec{r}\,';\tau)\;\Psi(\vec{r}\,',0)\; d^3\vec{r}' \qquad (11)$$

The function G which is usually called the **Green's function** (or the propagator) is responsible for time evolution of wavefunction $\Psi(\vec{r},\tau)$ by an arbitrary time τ and we can use the Green's function for finding the solution $\{\phi_0(r),E_0\}$. In [3,4] Kalos introduced a method which allows one to use Monte Carlo to sample from the exact propagator even when it is unknown. It is based on MC iteration of the Fredholm integral equation starting with

an exact propagator G_0 for simple potential V_0. The procedure described there is rather complicated and therefore an approximation is widely used.

Our propagator is given by the following formula

$$G(\vec{r},\vec{r}';\tau) = \langle \vec{r} | e^{-\tau(\hat{T} + (\hat{V}-E_T))} | \vec{r}' \rangle \tag{12}$$

If we assume short times τ (and we will depict it explicitly by writing $\Delta\tau$ instead of τ), we can make an approximation which is based on the fact that for small times the kinetic ($\hat{T}$) and the potential ($\hat{V}$) operators commute. Then we have

$$\begin{aligned} G(\vec{r},\vec{r}';\Delta\tau) \approx \langle \vec{r} | e^{-\Delta\tau\hat{V}/2} \; e^{-\Delta\tau\hat{T}} \; e^{-\Delta\tau\hat{V}/2} \; | \vec{r}' \rangle \; e^{\Delta\tau E_T} = \\ = w(\vec{r},\vec{r}';\Delta\tau)\, G_D(\vec{r},\vec{r}';\Delta\tau) \end{aligned} \tag{13}$$

where

$$w(\vec{r},\vec{r}';\Delta\tau) = e^{-\Delta\tau([V(\vec{r})+V(\vec{r}')]/2-E_T)} \tag{14}$$

and

$$\begin{aligned} G_D(\vec{r},\vec{r}';\Delta\tau) = \langle \vec{r} | e^{-\Delta\tau\hat{T}} | \vec{r}' \rangle = \\ = (4\pi D\Delta\tau)^{-3N/2} \, e^{(\vec{r}-\vec{r}')^2/(4D\Delta\tau)} \end{aligned} \tag{15}$$

where $D = 1/2$.

The w term is **the rate of decay of the particles** in the case when only the $\hat{V}$ operator is present and G_D is the **diffusion propagator** for the equation with the kinetic energy operator $\hat{T}$ only. Equation (6) is a diffusion equation in a 3N dimensional space and can be easily simulated [5]. In this equation $\Psi(\vec{r},t)$ plays the role of the density of diffusing particles. In case of having just the kinetic energy operator $\hat{T}$, equation (6) would be just an ordinary **diffusion differential equation** (with the diffusion coefficient 1/2) which can be easily simulated by a random walk of the ensemble of "walkers" through the given configuration space of the system. On the other hand if only $V(\vec{r})-E_T$ is present, we would have a kinetic **process of decay**

and/or creation of walkers depending on the sign of the $V(\vec{r})-E_T$ quantity. Thus the simulation process in the case when both parts of the operator $\hat{H}$ are present will be a combination of diffusion and decay/creation (called branching) of walkers as can be seen from the propagator (13).

If we want to treat the $\Psi(\vec{r},t)$ function as a density of walkers, it must be positive everywhere, which is the main problem in solving the SchE **for fermions** in this way. It is known that a fermion's wave function must be **antisymmetric under any exchange of coordinates** to obey the Fermi statistics. The above problem can be treated by the fixed-node approximation which deals with different regions of the wave function separately. We will come to this point later.

In general finding the lowest eigenstate ϕ_0 and eigenvalue E_0 is done by taking the limit $\tau\to\infty$ which is achieved theoretically by

$$\Psi(\vec{r},\tau\to\infty) = \int G(\vec{r}\ ,\vec{r}'\ ;\tau\to\infty)\Psi(\vec{r}',0)d^3\vec{r}' \qquad (16)$$

and achieved computationally by making a lot of iterations of the following kind

$$\Psi(\vec{r},\tau) = \int G_{ST}(\vec{r}\ ,\vec{r}'\ ;\Delta\tau)\Psi(\vec{r}',\tau-\Delta\tau)d^3\vec{r}' \qquad (17)$$

until all the higher excitations vanish (see eq.(8)). This is **equivalent** to moving the walkers throughout the configuration space by imposing random diffusion on them and making them decay or be created, thus simulating the process. There are well defined computational procedures for simulating (17). What we get at the end of such a procedure is an equilibrated ensemble with the density of walkers corresponding to the ground state wave function ϕ_0.

The other method, namely GFMC, uses a similar type of integral to achieve the same goal

$$\Psi^{(n+1)}(\mathbf{R}) = E\int G(\mathbf{R},\mathbf{R}')\Psi^{(n)}(\mathbf{R}')d^3\mathbf{R}' \qquad (18)$$

where

$$\hat{H}\, G(\mathbf{R},\mathbf{R}') = \delta(\mathbf{R}-\mathbf{R}') \tag{19}$$

The Green's functions are related in the following way

$$G(\mathbf{R},\mathbf{R}') = \langle \mathbf{R}| \frac{1}{\hat{H}} |\mathbf{R}'\rangle = \langle \mathbf{R}| \int_0^\infty e^{-\tau \hat{H}} d\tau \,|\mathbf{R}'\rangle = \int_0^\infty G(\mathbf{R},\mathbf{R}',\tau) d\tau \tag{20}$$

Although both procedures are exact, they cannot be used for computation, because the Green's function is not known exactly. Therefore an approximation must be used in both cases. As we have already mentioned, the **short time approximation** for the propagator is used in case (13), where we have assumed that the kinetic and the potential energy operators commute.

Now we will describe the DQMC **procedure** (which is basically the computer code algorithm), and highlight the problems which cause it to be inefficient around singularities of the potential energy $V(\vec{r})$. Then we show some techniques to overcome this inefficiency. In the DQMC procedure we start with a set of M configurations of $\vec{r}$ which we call the ensemble of walkers. It is better to start with these walkers distributed in regions where we expect the wave function to be large. We set the local time of each of these configurations equal to $\tau=0$ and

1) take one of the configurations and displace each of 3N coordinates randomly with a Gaussian distribution given by (15). This is the diffusion step. Then we

2) weight the configuration by $w(\vec{r},\vec{r}',\Delta\tau)$. The weighting is usually done by branching which means making duplicates if necessary, e.g. if $w\approx 1$ we leave the walker as it is (neither make any replicates nor delete it), but when $w\approx 4$ we make three extra identical configurations which will in principle evolve in different directions in the next step because of the randomness of the procedure. What is usually done computationaly in order to have integer number of walkers (n) is that we take $n=\mathrm{int}(w+\xi)$ which give us total number of walkers after weighting. The number ξ is a uniformly distributed random number ($0\le\xi<1$). If n=0 we terminate the walker by deleting the current configuration.

If we repeat the above two steps of diffusion and weighting for all M configurations from the ensemble then we have performed the integral (17). Our new ensemble now represents the wave function in a later time τ, so we must update the local time of all those configurations by $\Delta\tau$. As we have mentioned previously, we must adjust E_T during the procedure to be as close to the E_0 as possible, thus preventing the ensemble either from collapsing or from becoming too large. Once we have reached equilibrium for E_T (that means only statistical fluctuations are present) the configurations of the ensemble are samples of the ground state wave function.

Now concentrate for a moment on the branching part of the propagator, eq.(14), and realize what happens if the configuration (with position $\vec{r}$) approaches the singularity in the potential. The value of $V(\vec{r})$ will be huge and negative which will have the effect of producing too many duplicate configurations in the branching part of the process. Therefore the simulation will exhibit the **instability** not only because the next time step "integration" will need more CPU time to be completed because of more walkers, but their local energies will also bias the average energy we are computing and therefore increase the statistical uncertainties.

1.3. Importance sampling and the fixed-node approximation for fermions.

The problems with ineffective sampling due to potential energy singularities in DQMC were partly solved by Kalos [14] in 1974 when he suggested the **incorporation of the information** already known to improve the sampling efficiency and avoid singularities of the potential energy. The method is known as an **importance sampling scheme** and uses an analytical approximation to the wave function (WF) which we call the trial WF (Ψ_T). If we now express the imaginary time-dependent SchE (1.2.6) in terms of $f= \Phi\Psi_T$ rather than Ψ, we get a very similar equation to the previous one

$$\frac{\partial f(\vec{r},\tau)}{\partial \tau} = \frac{1}{2}\Delta f(\vec{r},\tau) - \nabla.(f(\vec{r},\tau)\, \vec{F}(\vec{r})) + (E_T - E_L(\vec{r}))\, f(\vec{r},\tau) \qquad (1)$$

where

$$E_L(\vec{r}) = \frac{H\Psi_T}{\Psi_T} \qquad (2)$$

is called the **local energy** which affects the branching and

$$\vec{F}(\vec{r}) = \frac{\nabla\Psi_T}{\Psi_T} \qquad (3)$$

is called the **drift velocity** and has an effect on sampling efficiency. There

are two differences between the equation we have just got and the previous one. The first is that we have an extra term

$$- \nabla.(f(\vec{r},\tau) \vec{F}(\vec{r})) \tag{4}$$

which is causing the drift of the walkers in such a way that it forces them to escape from the regions of small Ψ_T to those of large Ψ_T. Another difference is the branching term. In the propagator used for computer simulation it will appear as

$$w_{IS}(\vec{r},\vec{r}',\Delta\tau) = e^{-\Delta\tau\left(\frac{E_L(\vec{r}) + E_L(\vec{r}')}{2} - E_T\right)} \tag{5}$$

If we realize that the kinetic energy as well as the potential energy part are present in w_{IS}, we can see that the singularities of potential energy are partly eliminated by the kinetic energy operator. Although we do not get rid of the singularities completely (we are still having some because our trial WF is only an approximation to the exact WF), we avoid the most severe ones. What is very easy to notice is that in case of having the **exact** WF rather than a guess, whatever the position $\vec{r}$ is we have the same value for $E_L(\vec{r})$ because

$$\frac{H\Psi_T}{\Psi_T} = E_{exact} \tag{6}$$

If we then adjust $E_T = E_{exact}$ we get rid of the branching completely.

In the case when Ψ_T is a **good approximation** to Ψ_0 the variance of the E_L and therefore the statistical uncertainty will vanish giving us the exact energy. Even if we don't have the best possible trial function, after incorporating all the information we know about Ψ_0 into the estimate (e.g. cusp conditions for electrons) we get much better results than without importance sampling.

We note, however, that in order to interpret eq.(1.2.6) in terms of the diffusion process, the Ψ must be positive for the whole region because it represents the population density of walkers. It might seem that we are restricted to WFs without nodes, like Bose system WFs, and we can't treat Fermion systems by DQMC. Fortunately that's not true. The only thing which restricts us, however, is the requirement of having the nodes fixed during the simulation. If we therefore multiply the desired solution $\Phi(\vec{r},\tau)$ with the trial one $\Psi_T(\vec{r},\tau)$ and ask

$$f(\vec{r},\tau) = \Phi(\vec{r},\tau) \cdot \Psi_T(\vec{r},\tau) \qquad (7)$$

to be non-negative everywhere, than we can interpret it as a density of walkers in eq.(1). By doing that, we are restricting ourselves to the class of the functions Φ with exactly the same nodes as Ψ_T. The **fixed-node approximation** solves the SchE separately in different "sign" regions of WF.

If we know the **exact position** of the WF nodes of a given molecule, we could solve eq.(1) **exactly** (only subject to statistical fluctuations). To know the position of the nodes is, however, impossible, as it is essentially knowing in advance the exact wavefunction.

Using the fixed nodes does not violate the variational principle, because the expectation value of the energy calculated using $f(\vec{r},\tau)$ is still an upper bound to the exact ground state energy [15].

1.4. Optimization of Wave Function.

It was shown by Reynolds and Ceperley [15,16] that DQMC with importance sampling (or the fixed-node approximation in the case of fermions) is capable of providing results which are comparable with CI for small molecules like H_2 ,LiH ,Li_2 and H_2O. What is usually used for the trial WF is either the Hartree-Fock (HF)determinant or Multiconfiguration-Self-Consistent-Field WF (MCSCF). In addition to this an extra **Jastrow correlation function** (J) is usually present as a multiplicative factor to achieve good electron-electron cusp conditions as was mentioned previously.

It was shown in [15,16] that these WFs used in VQMC usually recover about 15%-80% of the correlation energy, but in the case of DQMC this is about 80%-100%. What one always wants when performing importance sampling using DQMC is a good trial WF. It is obvious that HF-J or MCSCF-J are usually not the best WFs to use. It is not very difficult to obtain better trial WFs at some extra cost of optimization. Different approaches can be used. Some of them try to minimize the expectation value of the energy by adjusting the parameters, some **minimize the variance** [17,18] of the local energy, namely

$$\sigma^2 = \frac{\sum_{i=1}^{M} \left(\frac{H\Psi_T}{\Psi_T} - E_g\right)^2 \left(\frac{\Psi(i)}{\Psi_0(i)}\right)^2}{\sum_{i=1}^{M} \left(\frac{\Psi(i)}{\Psi_0(i)}\right)^2} \qquad (1)$$

where E_g is a guess of the energy of the corresponding quantum state, Ψ_0 is the best possible WF before the optimization, usually HF-J and M is the product of the number of iterations N_{it} and the number of elements in the ensemble N_{ens} . Though the optimization procedure doesn't always find the global minimum (it might end in a local one because of difficulties always present in nonlinear optimizations), it's relatively easy to find the WF which is better in terms of statistical errors and the expectation values than the starting one. The advantage of the variance optimization is that the quantity we are optimizing has a known minimal value, namely zero. Moreover, it is valid not only for the ground state, but also for the excited states thus in principle allowing us to find excited states as well. All this is valid for molecules with small Z when by optimizing the wavefunction we **equally optimize** all the parameters which are present in the WF formula. What we mean by this is that since all the electrons are on the same energy scale, therefore their contributions to the total local energy are on the same energy scale and the fluctuations which determine the variance are on the same scale as well. This is not, however, valid for large Z molecules where the energies of the innermost electrons are much larger in their values (and fluctuations as well) so that their contribution to the overall variance is larger. We know that **valence electrons** are those which are **chemically important** and therefore to have "chemically good" WFs we have to use some other procedure which can provide us at least with some crude way of obtaining better parameters than we are able to get just by hand without using any optimization.

Let us now briefly explain the procedure which is usually used for optimizing the wavefunctions of small Z molecules by using the **analytical derivatives** of the variance of local energy [18]. Using analytical rather than numerical derivatives makes the algorithm more powerful, because instead of computing the local energy three times to get the given derivative of E_L due to one parameter we can have the derivative as a by-product of a single local energy evaluation.

We want to find parameters of Ψ_T which minimize the variance of the local energy

$$\sigma^2 = \frac{\sum_{i=1}^{M} (E_L(i) - E_g)^2 \, w(i)}{\sum_{i=1}^{M} w(i)} \qquad (2)$$

which is actually identical to (1). Quantities to be computed are:
1) Ψ_T , which is simply a product of Slater determinants multiplied by the Jastrow factor

$$\Psi_T = \Psi^{\uparrow} \Psi^{\downarrow} \Psi^{c} \qquad (3)$$

2) $E_L(i)$, which is the local energy. After splitting it into the contributions from spin up, spin down or correlation part denoted by superscripts $^{\uparrow}$, $^{\downarrow}$ or ^{c} respectively, we have the following expression where two Greek indices in the same term indicate the summation through that index.

$$E_L = -\frac{1}{2}\left(g^{\uparrow} + g^{\downarrow} + g^{c} + 2\left(\mathbf{F}^{\uparrow}{}_{\alpha} . \mathbf{F}^{\uparrow c}{}_{\alpha} + \mathbf{F}^{\downarrow}{}_{\beta} . \mathbf{F}^{\downarrow c}{}_{\beta} \right) \right) + U \qquad (4)$$

3) all the first derivatives of the above quantities.
We can easily get eq.(4) by plugging the expression for Ψ_T into eq.(1.3.2) and we find that the following holds:

$$g = \frac{\nabla^2 \Psi}{\Psi} \qquad (5)$$

$$\mathbf{F}_i = \frac{\nabla_i \Psi}{\Psi} \qquad (6)$$

where the arrows and the superscript denote the contribution from the different part of the wavefunction and U is the potential energy. Since all other necessary formulas are in [18] we will not go into more details, but still some points should be mentioned.

First thing is that we assume **quadratic** behavior of the variance with

respect to any parameter around the minimum and therefore we can use a simple quadratic estimate for the location of the minimum having the first and second derivative in that point.

$$p_j^{new} = p_j - \frac{(\frac{\partial \sigma^2}{\partial p_j})}{(\frac{\partial^2 \sigma^2}{\partial p_j^2})} \lambda^{-1} \tag{7}$$

Parameter λ is "guiding" the optimization. The above assumption can be used, however, even in the case of more complicated behavior. In the actual optimization, λ is usually around 10 which means that rather than making the full step towards the minimum in the given parameter direction, we step only a bit in the direction of the gradient. This method is usually called the **steepest-descent technique** and is widely used in other fields as well.

Optimization of a quantity like σ^2 which is obtained by sampling some trial wave function has a unique feature not present in other optimizations, and that is its **statistical fluctuations**. Since we can get only an estimate of this quantity with nonzero standard error, we do not have to compute its exact second derivative, but only an approximate one. This approximation is the standard approximation used with **least-squares** fits in other areas as well. The approximation is as follows

$$\frac{\partial^2 \sigma^2}{\partial p_j^2} = 2 \sum_i^M \{\frac{\partial}{\partial p_j} [\frac{(E_L(i) - E_g) w(i)}{(\sum_i w(i)^2)}]\}^{1/2} \tag{8}$$

This allows us to express the approximate second derivatives in terms of the first derivatives of the local energy and the wave function.

The other thing connected to the statistical fluctuations of the variance is that we can't reject all parameter changes which result in an increase of σ^2 . Only if the new σ^2 (the one for a new set of parameters) is 1.5 times larger than the "best"variance so far, do we backtrack to the best values of

parameters (and the variance) so far. Then we multiply the best σ^2 by a factor of 1.5 to prevent the procedure from being cornered later by the fluctuations (a value which was found empirically).

Another feature is the replacement of stale configurations by those which are moving and therefore represent the WF more appropriately. Since during the optimization procedure the parameters are often changed and therefore the nodes of the WF are moving as well, the configurations might be trapped in the vicinity of the node and under the **Metropolis sampling** would never move. By replacing these trapped configurations equilibration is reached faster.

Now we will describe Metropolis sampling. The calculation of expectation values is the essential part of the Monte Carlo methods. In order to have very precise results, we have to sample the space of all possible configurations. There are basically two possibilities. The first one is the **straightforward sampling** for which the configurations are chosen **uniformly**. If the function, the expectation value of which we are trying to find, is peaked around some point, then this straightforward sampling is ineffective and converges slowly. There are many examples of sharply peaked functions in physics. One of them is the canonical ensemble distribution function which is peaked around the average value of the energy of the system <E>. The second choice, which treats the mentioned disadvantage, is Metropolis sampling. In this case an appropriate measure must be chosen to **prefer** the configurations which give the dominant contribution to the integral in order to make the sampling more efficient. Specifically, the configurations are seen to be distributed according to an underlying probability distribution function (PDF), which in the case of statistical physics, is a priori unknown. Fortunately in the case of QMC we sample using a known WF. The **transition probability** for transition from one configuration to another in QMC is derived using this wave function. Ultimately the ensemble will be distributed according to the desired distribution. This can be shown using the **detailed balance** principle. Metropolis et.al [52] suggested simple rules for accepting or rejecting the configurations after an attempted move. This approach is a non-uniform sampling scheme which exhibits better convergence than the uniform one.

1.5. Large Z systems and critical slowing down effect.

As we have mentioned previously, the effect known as **critical slowing down** applies to large Z systems. This effect is present not only in QMC calculations, but in other large scale computer simulations as well, used for instance in phase transitions and cooperative phenomena [13]. Since phase transitions are caused by the effects of large size systems, to simulate them properly one needs the system (size L) to be as large as possible to eliminate the finite size effect to the acceptable level. In the case of spin systems, for instance, the time required for the set of spins to **lose it's coherence** (in the vicinity of the critical temperature) exhibits the following dependency on L :

$$T = L^2 \tag{1}$$

Therefore, sooner or later we have to face the problem of insufficient CPU performance, affecting the largest possible size of the system we can take into consideration.

The same holds for DQMC calculations of molecular structure. In this case the CPU time needed to reduce the the statistical uncertainties to the acceptable level of chemical accuracy depends on the nuclear charge of the atoms as follows

$$T = Z^{6.5} \tag{2}$$

The above dependency was estimated by Hammond et al. [11] and Ceperley [12]. There are two reasons for such steep dependency on Z. One of them arises because an increasingly large fraction of the total energy comes from the core rather than valence part which makes the energy fluctuations very large and therefore requires the simulation to be run for longer time to reduce the total error in computing the energy. The above mentioned reason comes from the fact that the core energy rises as Z^2 . There is yet another more substantial reason, and that is the **reduction of the phase**

space of the core electrons. They have to take smaller steps and therefore the simulation times $\Delta\tau$ have to be taken in accordance with the following formula

$$\Delta\tau = Z^{-2} \tag{3}$$

Because the phase space for the valence electrons is much larger than for core ones, valence electrons are unable to cover all the valence region and to follow the core electrons. Their sampling efficiency is therefore extremely bad.

Increasing the sampling efficiency for the valence electrons is the goal of many papers [11,12,19-25]. There are many methods treating this problem by using either pseudopotentials, local Hamiltonians or some other approaches. In the case of pseudopotentials the Z-dependence (2) is reduced to

$$T = (Z_{eff})^{3.4} \tag{4}$$

where the Z_{eff} is the screened nuclear charge. This approach, however, leads to the inaccuracies when there is significant electronic correlation between core and valence electrons. In that case all-electron calculations have to be used.

Let us briefly go through the methods which treat valence electrons and use different kinds of approximations for the core. In 1935 Hellman [26] and Gombas [27] for the first time published a paper where they introduced an effective potential which contained the core-valence repulsion and orthogonality condition. From that time the research goes basically in two directions. We can logically distinguish the following approaches. The first one (A) is **treating all electrons** and the second one (B) uses different kinds of **pseudo-operators** for the core electrons as well as relying on **core-valence partitioning of the wavefunction** $\Psi=\Psi_{core}\Psi_{val}$.

(A) One of the examples of this techniques is all-electron Effective Two-electron Potential (ETP) introduced by Carlson et al.[28]. In this method the two-electron potential for the core region is replaced by a one-electron

pseudopotential which reduces the variance in the core.
(B) There are three different approaches belonging to this group. The first method is the Effective Core Potential (ECP), suggested by Phillips and Kleinman [29] and improved later by many other researchers. The other method developed by Huzinaga et.al.[30,31] is called a Model Potential (MP). In both these methods the core-valence orthogonality condition is replaced by a **nonlocal pseudopotential**. The third method follows Bachelet's [32] suggestion to use the modification of the local momentum operator rather than the nonlocal potential energy operator. The method is called the Local Pseudo-Hamiltonian method (LPH). Both ECP and MP have the disadvantage of being non-local, and they have to be changed using approximations when implementing into the DQMC. The advantage of LPH is that it doesn't require any further approximations and can be used directly in the DQMC. There is yet one other method which is called the Damped-Core method (DC). It treats all the electrons, but simultaneously uses the wavefunction split into $\Psi=\Psi_{core}\Psi_{val}$. While simulating the core electrons according to $|\Psi_{core}|^2$ in a variational manner, it "solves" the Schrodinger equation for the valence electrons with Ψ_{val} as a guiding function, and making a branching of valence configurations in addition to diffusion and drift. In this DC method, suggested by Hammond, Reynolds and Lester [33], correlation effects which are not covered by pseudopotential techniques are taken into account.

This was a brief review of what is used in Valence Quantum Monte Carlo calculations in the present days. Up to date, however, no calculations including atoms heavier than fluorine were published using all-electron DQMC.

In our approach we want to present **the method of optimizing the valence part of the WF** of CuH molecule. We will then "check" the obtained WF by calculating spectroscopic properties of the molecule. Since these are very sensitive to the quality of the valence part of the WF, we can then make a statement about whether the method suggested here is picking up the correct parameters or not, in other words whether it is optimizing or not.

1.6. Transition metal hydrides - structure and bonding.

The **transition elements** are the elements that have atoms or ions with partially filled d orbitals, in other words they are the d-block elements. They are located in the middle of the periodic table. Transition elements have not only practical significance, since they provide the main structural metal (Fe), alloying and coinage metals (Ni,Mo,Co,Cr,Ag...) and the best metallic conductors (Cu,Au), but have also theoretical significance because of their ability to form complex ions and to catalyze chemical reactions. These elements are listed in the table 1.6.1 along with their atomic number, electronic configuration of the ground state and the group they belong to.

A transition element, by definition, must have a **partially filled d orbital** in either its atom or one or more of its ions. By applying this definition, we see that copper (group 1B) does not have d-orbital vacancies either in its atom or in Cu^{+}, but it has vacancy in the Cu^{2+}. On the other side zinc (group 2B) does not have d-orbital vacancy in either the free atom or in its ion. Therefore Zn doesn't belong to the transition elements. Zn,Cd,Hg are called representative metals.

Transition elements, often called **transition metals** (TM), form compounds with a wide range of oxidation numbers. Moreover they have a tendency to form a complex ion by surrounding the metal ion with ligands. These are called coordination compounds. It is known that certain TMs such as Cu,Y,La are important constituents of baked ceramic materials that exhibit zero electrical resistance under the critical temperature, which is at the present time on the order of 125K. There is no doubt that a better understanding of the structure and bonding of different compounds that contain TMs can have stimulating effect on the physics of superconductors and help to explain the mechanism of the high T_c superconductivity. To understand the nature of the chemical bond in a complex system composed

of TMs we must first have an understanding of what is happening in simple molecules.

Transition metal hydrides (TMH) and their positive ions are the first to be studied. They have a particularly fundamental importance in catalytic processes, either heterogeneous at metal surfaces or homogenous with different organometallic compounds, with special emphasis on activation of the hydrocarbons. There is, however, **no systematic experimental data** on the neutral TMHs, therefore in some cases theoretical values of different quantities are the only values available.On the other hand for positive ions of TMHs the **guided ion beam** approach provides us with accurate dissociation energies. This is of use not only to the study of the positive ions but can give us a calibration of the theoretical dissociation energies for neutrals.

From a theoretical point of view the quantities we are interested in are spectroscopic constants like (R_e,ω_e,$\omega_e x_e$), dissociation energy (D_e), separation between states (T_e) and the electric dipole moment (μ). The dipole moment is a very sensitive quantity in terms of wave function, thus providing us with a sensitive test of the wave function, and giving us insight into the bonding mechanism.

As far as the computational methods are concerned, we can start with the single-reference-based techniques. These techniques are economical and they have been used for TM systems as well as for lower atomic number elements. Because the TMs exhibit **quasidegeneracies** the **coupled-pair-functional** (CPF) method is not very useful because it is less stable to the quasidegeneracies. That is the reason why another **modified-coupled-pair-functional** (MCPF) method was developed [34] to deal with these quasidegeneracies. It has one other advantage over CPF which is the possibility to check the wave function quality by judging the percentage of the reference configuration in the wave function. MCPF is generally much better than the other more commonly used correlation method: **singles-and-doubles-configuration-interaction** (SDCI). The alternative to MCPF method is **coupled-cluster-singles-and-doubles** (CCSD) [35]. Because of above mentioned near degeneracies of TMs Moller-Plesset to second order (MP2) perturbation method doesn't work well. In general single-reference-based techniques work better only for those TM systems

where all d-electrons create bonds. We can apply single-reference-based techniques to the TMs on the both sides of the row but not in the center. Those TMs in the middle have many open-shell d-electrons which create many low lying states and their mixing needs to be described by multi-reference functions. Not to forget multi-reference-based techniques we have to mention the **complete-active-space self-consistent-field** (CASSCF) method in combination with **multi-reference-configuration-interaction** (MRCI) or with the **average-coupled-pair-functional** (ACPF) method [36]. The first part of the method (CASSCF) in both cases deals with the metal quasidegeneracies while the second part (MRCI or ACPF) takes care of extensive correlation computations. The energy in CASSCF-MRCI is comparable to full configuration interaction (FCI) for elements with small atomic numbers, even using small basis sets, but it does not give accurate dipole moments for TM molecules due to the steep growth of the expansion with the number of active orbitals. We have to account only for the most important degeneracies. This introduces an orbital bias. On the other hand ACPF can account for higher excitations and reduce the above mentioned bias. Both methods, with the help of the **iterative-natural-orbital** procedure [37] (INO), effectively reducing the orbital bias, usually yield accurate results, but their disadvantage is that they are expensive for routine calculations.

A common feature in the bonding in either TMHs or their positive ions is that the **bond involves the mixture of different atomic states.** Therefore it is of enormous importance to compute the atomic separations with sufficient precision before trying to proceed further. In general using double or triple zeta one-particle basis functions with the polarization function we can have sufficiently good separation between $d^n s^2$ and $d^{n+1} s^1$ electronic configurations using methods like CASSCF-MRCI, MCPF or QCISD(T). There are several theoretical papers [38,39] on neutral TMHs calculating the spectroscopic constants (R_e,D_e,ω_e) of the ground state and describing the nature of chemical bond in these systems. We can not, however, while having all these values for the first row systems, generalize it to the second row transition metal hydrides and their positive ions because there are still some differences, mainly due to the different atomic asymptotes and their separations. There are a couple of

examples. While ScH and YH are both formed by the same sd hybridization thus creating the $^1\Sigma^+$ ground state, the TiH and ZrH on the other hand do not exhibit a similarity in the bonding type. While TiH bonding is composed of the mixture of $^3F(3d^24s^2)$ and $^5F(3d^34s^1)$ atomic states, the ZrH bond rests on the formation of sd hybrids from the $^3F(4d^25s^2)$ states. We can see this in the table of Mulliken d-population. The second example is MnH and TcH. The ground state of MnH ($^7\Sigma^+$) is derived from the $^6S(3d^54s^2)$ state of Mn only because the nearest upper state $^5D(3d^64s^1)$ doesn't contribute because of it's high energy. In the case of TcH this ^5D state is much lower in energy and mixes into the wave function thus creating $^5\Sigma^+$ ground state.

It is a well known that the electric dipole moment is a very good measure of the quality of the wave function. It is sensitive to the mixing of atomic asymptotes in the bonding orbital [38]. There are basically two kinds of atomic asymptotes in the transition metals. One of them $\mathbf{d^{n+1}s^1}$, which forms an **s-s** bond between metal and hydrogen is polarized towards H and therefore gives large dipole moment. This is the case of CrH and its asymptote $3d^54s^1$. The other one is a $\mathbf{d^ns^2}$ asymptote, which forms bonds by **sp** hybrid orbital formation. The bonding orbital of two hybrid orbitals is polarized toward H and the antibonding orbital away from H. The result is a smaller electric dipole moment in MnH (asymptote $3d^54s^2$), see table 1.6.2. The magnitude of the dipole moment reflects the relative mixture of the two atomic asymptotes. For the pure $3d^n4s^2$ the dipole moment is small and for pure $3d^{n+1}4s^1$ it is large. Thus CuH has the largest μ along with CrH in the first row and AgH and MoH in the second row (table 1.6.2).

In table 1.6.3 there are some results for CuH for three different computational methods, along with the available experimental values [40]. The ground state of copper hydride ($^1\Sigma^+$) is dominated by the $3d^{10}4s^1$ electronic configuration, though CPF method indicate presence of a $3d^94s^2$ contribution. Both CPF and MCPF methods give very similar results. On the other hand, despite the fact that not only the energy but also spectroscopic parameters (R_e,ω_e,D_e) are very close for different methods, they differ a lot from experiment.

Investigation using larger basis set at the CPF level with [9s7p4d

3f1g/4s3p2d] GTO reduced R_e only by 0.028, but if the Darwin and mass velocity relativistic contributions are included as perturbations of the first order, reduction of R_e was 0.05 a.u. Similar behavior can be seen in ω_e (increasing by 78 cm^{-1}) and D_e (increasing by 0.14 eV). As far as spectroscopy of CuH is concerned [41] one must include spin-orbit coupling to explain the spectra theoretically.

In general, theoretical studies have shown that the bonding in TMHs arises from $3d^n4s^2$ and $3d^{n+1}4s^1$ (for the first row TMs) and is correlated with the separation between these atomic asymptotes. The $3d^n4s^2$ bonding forms generally two **sp** hybrid orbitals, one pointing toward hydrogen and the other one at the opposite side. The $3d^{n+1}4s^1$ occupation forms a **4s metal 1s H bond** polarized toward H. Bonding is influenced by the separation between these asymptotes and can be a mixture of both these occupations.

Our approach is quite different from those explained above. Rather than having complicated mixtures of different occupations, we rely on the optimized wave function for just one occupation.

TABLE 1.6.1: Transition Metals.

Symbols used in table : [Ar] = $1s^2 2s^2 2p^6 3s^2 3p^6$

[Kr] = $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6$

[Xe] = $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^{10} 5p^6$

at.number	symbol	el. configuration	group
21	Sc	[Ar]$4s^2 3d^1$	3B
22	Ti	[Ar]$4s^2 3d^2$	4B
23	V	[Ar]$4s^2 3d^3$	5B
24	Cr	[Ar]$4s^1 3d^5$	6B
25	Mn	[Ar]$4s^2 3d^5$	7B
26	Fe	[Ar]$4s^2 3d^6$	8B
27	Co	[Ar]$4s^2 3d^7$	8B
28	Ni	[Ar]$4s^2 3d^8$	8B
29	Cu	[Ar]$4s^1 3d^{10}$	1B
39	Y	[Kr]$5s^2 4d^1$	3B
40	Zr	[Kr]$5s^2 4d^2$	4B
41	Nb	[Kr]$5s^1 4d^4$	5B
42	Mo	[Kr]$5s^1 4d^5$	6B
43	Tc	[Kr]$5s^2 4d^5$	7B
44	Ru	[Kr]$5s^1 4d^7$	8B
45	Rh	[Kr]$5s^1 4d^8$	8B
46	Pd	[Kr]$4d^{10}$	8B
47	Ag	[Kr]$5s^1 4d^{10}$	1B
57	La	[Xe]$6s^2 5d^1$	3B
72	Hf	[Xe]$6s^2 4f^{14} 5d^2$	4B
73	Ta	[Xe]$6s^2 4f^{14} 5d^3$	5B
74	W	[Xe]$6s^2 4f^{14} 5d^4$	6B
75	Re	[Xe]$6s^2 4f^{14} 5d^5$	7B
76	Os	[Xe]$6s^2 4f^{14} 5d^6$	8B
77	Ir	[Xe]$6s^2 4f^{14} 5d^7$	8B
78	Pt	[Xe]$6s^1 4f^{14} 5d^9$	8B
79	Au	[Xe]$6s^1 4f^{14} 5d^{10}$	1B

TABLE 1.6.2: Transition Metal Hydrides and their Ions[1].

first row	Re (a.u.)	ω_e (cm-1)	D_e (eV)	μ (D)	d popul
ScH(ScH^+)	3.390(3.457)	1587	2.27	1.37	0.84(1.29)
TiH(TiH^+)	3.440(3.289)	1548	2.06	2.19	2.30(2.34)
VH (VH^+)	3.249(3.143)	1635	2.33	2.02	3.40(3.37)
CrH(CrH^+)	3.201(3.031)	1647	2.13	3.81	4.83(4.45)
MnH(MnH^+)	3.313(3.122)	1530	1.67	1.24	5.05(5.10)
FeH(FeH^+)	2.973(3.029)	1915	1.67	2.90	6.52(6.19)
CoH(CoH^+)	2.895(2.923)	1842	1.94	2.74	7.60(7.22)
NiH(NiH^+)	2.807(2.810)	1987	2.69	2.56	8.65(8.26)
CuH(CuH^+)	2.851(2.730)	1852	2.63	2.95	9.80(9.45)

second row	Re (a.u.)	ω_e (cm-1)	D_e (eV)	μ (D)	d popul
YH(YH^+)	3.706(3.576)	1559	2.95	1.54	0.74(0.80)
ZrH(ZrH^+)	3.509(3.459)	1483	2.45	1.23	1.90(1.97)
NbH(NbH^+)	3.384(3.295)	1583	2.60	2.45	3.69(3.57)
MoH(MoH^+)	3.300(3.174)	1642	2.19	3.03	4.89(4.72)
TcH(TcH^+)	3.158(3.198)	1797	1.95	2.15	5.35(5.18)
RuH(RuH^+)	3.114(3.158)	1801	2.34	2.70	7.00(6.35)
RhH(RhH^+)	2.976(2.864)	2057	2.81	2.24	8.10(7.99)
PdH(PdH^+)	2.911(2.817)	1958	2.22	2.03	9.24(9.03)
AgH(AgH^+)	3.130(3.442)	1703	2.22	2.95	9.80(9.83)

[1] Ref. [53]
Transition Metal Hydrides - Structure and Bonding.
by C.W.Bauschlicher

TABLE 1.6.3: Theoretical and Experimental Results for Ground State of CuH ($^1\Sigma^+$)

method	E(a.u.) (a.u.)	Re (a.u.)	ω_e (cm-1)	De (eV)	μ (D)
SDCI[2]	-1639.80470	2.853	1815	2.29	3.880
CPF[2]	-1639.84003	2.852	1867	2.64	2.749
MCPF[2]	-1639.84076	2.851	1852	2.63	2.951
Expt.[3]	~~~	2.764	1941	2.85	~~~

[2]Ref. [40]
Theoretical dipole moments for the first-row transition metal hydrides. by Chong et.al

[33]Ref. [51]
K.P.Huber and G.Herzberg : Molecular Spectra and molecular Structure. (Van Nostrand Reinhold,New York,1979)

2.1. Splitting of the variance.

In the previous chapter we have briefly mentioned some methods, almost all of which are routinely used for the computation of the properties of transition metal hydrides. They can be used for other molecules as well. These methods extensively use correlation and all of them have very large basis sets corresponding to a huge space of functions, i.e. different possible contributing configurations coming out of the set of occupied and unoccupied molecular orbitals. This makes it a computationaly demanding task to calculate properties.

Our approach is substantially different from all those analytic techniques mentioned in the previous chapter. The difference is not only in the method we use, namely QMC, but also in the approach used for getting better properties of the system of interest, which is presently the copper hydride molecule (CuH). Rather than having a large basis space utilizing all of the possible contributing configurations we try to rely on a very simple wave function which is optimized in a certain sense.

In this work our aim was **to investigate the possibility of optimizing the valence part of the CuH wave function** . We will show by comparing to the more sophisticated techniques that a quite simple wave function is capable of providing us with fairly good properties. We will present a method used for the optimization as well as some parameters obtained using that method along with the calculated properties for the resulting wave functions.

Copper hydride is a large Z system containing a nuclear charge of Z_{Cu}=29. We know what the limitations are which apply to the large Z systems (chapter 1.5.). From the start it is obvious that we can't reach the precision of the small Z system calculations. We know that it is very difficult to calculate the value of the total energy of the system because there are large fluctuations in the energy. The electrons which contribute most are obviously the core electrons because of their very large negative energies. It is impossible to eliminate their energy fluctuations. The more

precise value of the energy we want to achieve (less statistical error), the larger sample we have to have. But for the computer simulation approach the large sample means larger CPU time. Previous estimates of the CPU time dependency on the nuclear charge are $T=Z^{6.5}$ [11,12]. This means : for a molecule with twice the nuclear charge as previously treated we have to have 90 times more CPU time to get results with comparable chemical accuracy.

The above mentioned dependency of CPU time incorporates not only the effect of large energy fluctuations, but also the differences of size of configuration space for core and valence electrons. Because we use the short time approximation for the Green's function, we use the same $\Delta\tau$ simulation steps for all the electrons. To achieve good sampling for energetically important electrons, we have to use the time steps which allow that. These time steps, however, are not large enough for the valence electrons to result in a good sampling. Some approximation has to be used to avoid this problem.

The approximation we have used in this work is based on assigning **different time steps to different electrons**. We call it the **split-tau** approximation. This assignment is done on the shell basis. The closer the electron is to the nucleus, the smaller is the $\Delta\tau$ value, but those electrons belonging to the same shell have the same value of the $\Delta\tau$. This will allow us to sample the core and valence regions **equally efficiently**. Since we are using the Variational Quantum Monte Carlo with the help of the Metropolis sampling procedure, this split-tau is not going to affect the result providing the electrons don't change places between shells. However, this is not quite true. Since the electrons are moved during diffusion and drift (for importance sampling), they exchange their shells from time to time, and by this exchange they change their corresponding $\Delta\tau$'s. We **assume** that this change of the electrons is not happening too often and disappears in the limit $\Delta\tau\rightarrow 0$, so that it will not affect the simulation too much. Later we will show that frequency of exchange is indeed acceptably low.

We know from chemistry that only a few electrons usually participate in the bond of the molecule. These are called **valence electrons**. The **core electrons**, which are closer to the nuclei, are strongly attracted and therefore can't freely change their positions in the molecule. Since the

valence electrons are not attracted to any particular nucleus, they are not confined to any particular place in the molecule. The molecular orbitals which describe their behaviour are usually spread out over very large regions in the molecule.

We must admit that the calculation of the total energy of the CuH system with the precision needed for the chemist would be impossible with today's computational resources using rigorous QMC techniques. There are, however, other properties chemists are interested in. Although they are energy-based properties (usually derivatives due to the bond distance Λ, and quantities derived from these derivatives), they exhibit completely different behaviour and can be simulated much more easily. The reason for that is obvious. Suppose we have the molecule CuH which is one of the simplest molecules in terms of geometry. We need only one parameter Λ to determine it's geometry. Now let us change the relative position of the hydrogen nucleus against Cu by changing Λ. It is obvious that the core electrons of the molecule, which are so close to the Cu will not feel the hydrogen changing it's position, because they are affected principally by the large Z_{Cu} charge. They are confined to the small region around Cu and the change of bond distance will have negligible effect on them. Therefore we do not expect large contributions from core electrons to the energy-based differential properties of CuH. This intuitive argument will be later supported by a more precise one along with the actual contributions to $\partial E/\partial \Lambda$ from those two groups of electrons.

In this work we want to find some relationship between σ^2_{val} which we call valence variance (it will be defined later in this chapter) and some energy-based molecular properties. In other words: is there any connection between the quality of the valence part of the wave function which consequently determines the properties and the value of the valence variance? What we have found allows us to tell what the region of the parameters should be that could possibly give us the best available parameters for the valence part of WF. For "further investigation" we have to use other criterion to exclude those parameters which are not good.

The first derivative of the ground state energy is the most important criterion. The reason is obvious. Suppose we use the experimental bond distance (as we actually do). We take the wave function and calculate the

first derivative $\partial E/\partial \Lambda$. Should it turn out to be far off the estimated zero value, the WF "doesn't fit" the molecule because it wants to **stretch it** more (when $\partial E/\partial \Lambda < 0$) or to **contract it** (when $\partial E/\partial \Lambda > 0$). This is contrary to the fact that we have used experimental geometry. The conclusion would be : the WF is not good.

Let us now explain the approach we have used in this work. Since we want to get rid of large core fluctuations contaminating the variance defined by

$$\sigma^2 = \frac{1}{M} \sum_{i=1}^{M} (E_L(i) - E_g)^2 \qquad (1)$$

we decided to split this variance into contributions from different shells as follows

$$\sigma^2_{split} = \sum_{s=1}^{N_{shell}} \left[\frac{1}{M} \sum_{i=1}^{M} (E_{L,s}(i) - E_{g,s})^2 \right] \qquad (2)$$

where i denotes the iteration, M is the number of iterations per block and N_{shell} is the number of shells (there is a certain arbitrariness of shell definition involved) in the system. Since in these simulations a **serial correlation** is present (our successive iterations are not totally independent), we group the iterations into the blocks, compute the block averages of them and only after that procedure is done can we look at block averages as independent samples.

As far as other quantities in formulas (1) and (2) are concerned, $E_L(i)$ is the average local energy for the current iteration i taken from the whole ensemble

$$E_L(i) = \frac{1}{NS} \sum_{j=1}^{NS} \sum_{k=1}^{N_{el}} E_{Lk}(j,i) \qquad (3)$$

and

$$E_{L,s}(i) = \frac{1}{NS} \sum_{j=1}^{NS} \sum_{k}^{\text{shell s}} E_{Lk}(j,i) \qquad (4)$$

NS is the number of elements in the ensemble, N_{el} is the number of electrons. Quantity E_g (in eq.(1)) is an estimate of the ground state energy eigenvalue. In the "standard" optimization (eq.1), we can adjust the optimized parameters so that the local energy will give us values very close to the E_g for each iteration and the value of σ^2 will become small. In the limit of exact WF we get $\sigma^2=0$.

In the case of **split variance** (eq.(2)) this is however not so. In this case $E_{L,s}(i)$ is the average local energy for those electrons which belong to the shell s . If the total system could be divided into separate **independent** (noninteracting) subsystems, then we could speak of the energy eigenvalues for separate systems and we could insert these values (if we knew them) into $E_{g,s}$ and proceed exactly as in the standard optimization using formula (2) instead of (1). However, this separation is in principle not possible. Since the electrons in the system are interacting with each other by Coulomb interaction we can't therefore find parts of the total system which are independent of each other. The only reasonable thing we can do is to take $E_{g,s}$ equal to the block average of the $E_{L,s}(i)$ quantity and hope that the core contributions to this $E_{L,s}(i)$ (for s=valence) will somehow "smooth out" each other during the averaging. If the valence part of the WF is not appropriate, we can expect larger fluctuations of $E_{L,\text{valence}}(i)$ and therefore

$$\sigma^2_{\text{valence}} = \frac{1}{M} \sum_{i=1}^{M} (E_{L,\text{valence}}(i) - \langle E_{L,\text{valence}} \rangle)^2 \qquad (5)$$

where

$$<E_{L,valence}> = \frac{1}{M}\sum_{i=1}^{M}(E_{L,valence}(i)) \qquad (6)$$

will have larger value. Because we can't judge anything from just one such $\sigma^2_{valence}$, we use many of these values computed from successive blocks of iterations. Details of all this will be mentioned in the next part of this work.

Finally we define our valence energy. The shell split of the rest of the system is quite arbitrary because we will not use other shell variances anyway. We base our method **only** on the valence variance.

The total Hamiltonian of the system (as in 1.2.1.) consists of the kinetic energy part (which is perfectly separable into contributions from different electrons) and the two potential energy parts. One of them is due to the nucleus-electron interaction and can be separated as well, but for the electron-electron interaction it can be done only approximately. We have done this on a half-half basis which means we assign one half of the Coulomb energy

$$\frac{1}{|\vec{r}_i - \vec{r}_j|}$$

to electron "i" and the second half to electron "j", no matter which shell it belongs to. In our case we have two valence electrons . They have unlike spins and they are the two most distant electrons (based on the distance from copper nucleus). Because of the diffusion and drift used in the simulation, we have to reassign the electrons into shells after each move since they change their distances from Cu and thus change the shells to which they belong. A procedure which ensures the "switching" of the electrons was incorporated into the program. Let us now denote the two outermost (valence) electrons by the labels n' and $\underline{n'}$ for spin up and spin down respectively. The valence energy operator in atomic units then will be

$$H_{valence} = -\frac{1}{2}\left(\Delta_{n'} + \Delta_{\underline{n'}}\right) - \sum_{\alpha} \frac{Z_\alpha}{|\vec{r}_{n'} - \mathbf{R}_\alpha|} - \sum_{\alpha} \frac{Z_\alpha}{|\vec{r}_{\underline{n'}} - \mathbf{R}_\alpha|} + \frac{1}{2} \sum_{(i \neq n')}^{N_{el}} \frac{1}{|\vec{r}_i - \vec{r}_{n'}|} + \frac{1}{2} \sum_{(i \neq \underline{n'})}^{N_{el}} \frac{1}{|\vec{r}_i - \vec{r}_{\underline{n'}}|} \qquad (7)$$

The nuclear-nuclear potential energy was not included since for the fixed bond distance it represents only an additive constant which does not change during the simulation.

The valence local energy is computed in a similar way as it is for the total Hamiltonian local energy

$$E_{L,valence} = \frac{H_{valence}\Psi_T}{\Psi_T} \qquad (8)$$

By this we get only the contributions to the local energy from the valence electrons.

2.2. Wave function and basis set used.

It is obvious as was pointed out in Chapter 1.3. on importance sampling and fixed node approximation, that the wave function plays an important role in QMC calculations. Both DQMC (which solves the Schroedinger equation), and VQMC (which estimates the quantities variationally) need a good trial wave function for good performance. The trial WF not only guides the walkers preferentially into the regions of configuration space where $|\Psi_T|^2$ is large so that it is sampled correctly, but makes the branching (in DQMC) smoother as well. In our case of VQMC the thing which is important is the nodal structure of Ψ_T. The more accurately the nodes are determined, the more precise will be the final answer. Since VQMC moves the walkers using the diffusion and drift without branching and without deleting them after crossing the nodes, after a long time the ensemble forms a discrete representation of the density $f(\vec{r},\infty) = |\Psi_T(\vec{r})|^2$. The expectation value of the energy is

$$\frac{\int f(\vec{r},\infty)\,\frac{H\Psi_T}{\Psi_T}\,d^3\vec{r}}{\int f(\vec{r},\infty)\,d^3\vec{r}} = \frac{<\Psi_T|H|\Psi_T>}{<\Psi_T|\Psi_T>} \qquad (1)$$

The choice of the trial wave function is to a great extent a balance between two factors that are somewhat exclusive each other. Those two factors are : incorporating enough terms into the form of WF to be able to describe different effects which are present in the system with an appropriate accuracy on one side, and the compactness and computational convenience which will pay off by allowing us to have a larger ensemble for sampling on the other side. In this work we choose to have the most compact wavefunctions possible.

The usual form of the trial WF in all kinds of QMC calculations is

$$\Psi_T = \Psi^{\uparrow}\,\Psi^{\downarrow}\,\Psi^{c} \tag{2}$$

which basically consists of two parts. The first is the product of two Slater determinants for the electrons with different spins and the second is the Jastrow correlation factor [42]. Sometimes the first part can be taken to be more general

$$\sum_{k=1}^{N_k} d_k\,\Psi^{\uparrow}_k \Psi^{\downarrow}_k \tag{3}$$

which reminds us of the configuration interaction approach accounting for different possible electron configurations in the system. Since the Ψ^c is symmetric for any exchange of electrons, the form given by eq. (2) provides us with the WF which is antisymmetric for interchange of spinlike electrons. This corresponds to the column change in one of the Slater determinants $\Psi^{\uparrow(\downarrow)}$.

More specifically, for a closed shell system, the determinants $\Psi^{\uparrow}(\vec{r}_1,\ldots,\vec{r}_n)$ and $\Psi^{\downarrow}(\vec{r}_{n+1},\ldots,\vec{r}_N)$ have the form

$$\begin{vmatrix} \phi_1^{\alpha}(\vec{r}_1) & \cdots & \phi_1^{\alpha}(\vec{r}_n) \\ \vdots & \ddots & \vdots \\ \phi_n^{\alpha}(\vec{r}_1) & \cdots & \phi_n^{\alpha}(\vec{r}_n) \end{vmatrix} \tag{4}$$

and

$$\begin{vmatrix} \phi_1^{\beta}(\vec{r}_{n+1}) & \cdots & \phi_1^{\beta}(\vec{r}_N) \\ \vdots & \ddots & \vdots \\ \phi_n^{\beta}(\vec{r}_{n+1}) & \cdots & \phi_n^{\beta}(\vec{r}_N) \end{vmatrix}$$

respectively. Here N is the number of electrons and n = N/2. Positions $\vec{r}_1$ through $\vec{r}_n$ denote the spin-up electrons and $\vec{r}_{n+1}$ through $\vec{r}_N$ spin-down electrons. Functions ϕ^{α}_i or ϕ^{β}_i are called **molecular orbitals** (MO) as is usual in the Hartree-Fock (HF) and other approaches. In the closed shell system, $\phi^{\alpha}_i(\vec{r})$ and $\phi^{\beta}_i(\vec{r})$ molecular orbitals are therefore identical. Two electrons occupy one orbital which is not contrary to the Pauli principle. These molecular orbitals consist of a linear combination of **atomic orbitals** (AO) in the following form.

$$\phi_i(\vec{r}) = \sum_{\nu=1}^{N_{basis}} C_{i\nu}\, \chi_\nu(\vec{r}) \qquad (5)$$

Quality of the basis as well as the coefficients for linear combination of AO ("MO coefficients") are the other very important factors when dealing with the wave functions.

The second part of the WF is the Jastrow correlation factor most simply written in the form

$$\Psi^c = \exp\left(\sum_{i<j} U_{ij} \right) \qquad (6)$$

where

$$U_{ij} = \frac{a|\vec{r}_i - \vec{r}_j|}{1+b|\vec{r}_i - \vec{r}_j|} \qquad (7)$$

Parameter a is determined by the requirement for the WF to have correct **cusp conditions** as two electrons approach each other where the

singularity in the potential energy is cancelled by the singularity in the kinetic energy . It has a value of 1/2 for unlike spin electrons and 1/4 for the like spin electrons. This term is always positive, therefore it does not change the nodes of the WF (completely determined by the determinants). The main reason why the Jastrow factor is important is the **variance reduction**. Although it does not affect the fixed-node (DQMC) energy, it affects the variational one.

Slater determinants composed of molecular orbitals have extremely high importance for the quality of the WF. These can be taken directly from the Hartree-Fock self-consistent field calculations, as is usually done, or even optimized further by VQMC technique like the one described in chapter 1.4. to better fit the QMC calculations.

What we have used in this work was the **single zeta** basis set of **Slater type atomic orbitals** [43] for neutral Cu atom obtained by Clementi and Roetti [44] . In addition to that, we used 1s and 2p0 orbitals with the centers on H for better description of the valence part of the WF. Slater type atomic orbitals (STO) are functions which are widely used for the calculations. Their form is

$$\chi_v^{STO}(\vec{r}) = N\, r^{n-1}\, Y_l^m(\theta,\phi)\; e^{-\xi r} \tag{8}$$

where n,l,m are quantum numbers, $Y_l^m(\theta,\phi)$ are normalized (non-complex) spherical harmonics

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}}\; P_l^{|m|}(\cos\theta) \begin{cases} \sqrt{2}\sin(|m|\varphi) & ;m<0 \\ 1 & ;m=0 \\ \sqrt{2}\cos(|m|\varphi) & ;m>0 \end{cases} \tag{9}$$

$P_l^{|m|}$ are associated Legendre functions, and N is the normalization factor

$$N = \frac{(2\xi)^{n+1/2}}{((2n)!)^{1/2}} \tag{10}$$

These orbitals remind us of hydrogen-like orbitals, but they have different

r-dependent polynomial parts.

All atomic orbitals in use are listed in the table 2.2.1. along with Clementi's exponents taken from [44] . We used exponent ξ=1. for the 2p0 AO instead of ξ=0.5 because we did not intend to build the shell structure on the hydrogen, only to introduce some perturbation into the 1s AO already present. The explicit formulas for STO are in the table 2.2.3 along with their normalizing factors. The reason why we have chosen STO basis set instead of some other one (e.g. Gaussian orbitals) is because STO describe electrons in atoms and molecules much better, and since we do not compute any integrals of the AO (as in the case of HF), there is no advantage to use GTO's. We could, however, use atomic hydrogen-like orbitals with "correct" nodal structure for the radial parts, but this turned out not to have any significant effect on the simulation of CuH molecule (this was also the case for other systems optimized before [18]).

Table 2.2.1.

TYPE	QUANT.NUMBERS			CENTRE	ξ
	n	l	m		
1s	1	0	0	Cu	28.3288
2s	2	0	0	Cu	10.5338
$2p_{-1}$	2	1	- 1	Cu	12.5541
$2p_0$	2	1	0	Cu	12.5541
$2p_{+1}$	2	1	+1	Cu	12.5541
3s	3	0	0	Cu	5.15648
$3p_{-1}$	3	1	- 1	Cu	4.87625
$3p_0$	3	1	0	Cu	4.87625
$3p_{+1}$	3	1	+1	Cu	4.87625
$3d_{-2}$	3	2	- 2	Cu	4.20186
$3d_{-1}$	3	2	- 1	Cu	4.20186
$3d_0$	3	2	0	Cu	4.20186
$3d_{+1}$	3	2	+1	Cu	4.20186
$3d_{+2}$	3	2	+2	Cu	4.20186
4s	4	0	0	Cu	1.20818
1s	1	0	0	H	1.0
$2p_0$	2	1	0	H	1.0

Table 2.2.2.
Associated Legendre polynomials.

Quantum numbers		$P_l^{\|m\|}(w)$	$P_l^{\|m\|}(\cos\theta)$
l	|m|		
0	0	1	1
1	0	w	$\cos\theta$
1	1	$\sqrt{1-w^2}$	$\sin\theta$
2	0	$(3w^2-1)/2$	$(3\cos^2\theta-1)/2$
2	1	$3w\sqrt{1-w^2}$	$3\cos\theta\sin\theta$
2	2	$3-3w^2$	$3\sin2\theta$

Table 2.2.3. Explicit form of the AO basis set functions.

$$\chi_{1,0,0}^{STO}(\vec{r}) = \sqrt{\frac{\xi^3}{\pi}}\, e^{-\xi r}$$

$$\chi_{2,0,0}^{STO}(\vec{r}) = \frac{1}{\sqrt{3}}\sqrt{\frac{\xi^5}{\pi}}\, r e^{-\xi r}$$

$$\chi_{2,1,-1}^{STO}(\vec{r}) = \sqrt{\frac{\xi^5}{\pi}}\, r e^{-\xi r} \sin\theta \sin\phi$$

$$\chi_{2,1,0}^{STO}(\vec{r}) = \sqrt{\frac{\xi^5}{\pi}}\, r e^{-\xi r} \cos\theta$$

$$\chi_{2,1,+1}^{STO}(\vec{r}) = \sqrt{\frac{\xi^5}{\pi}}\, r e^{-\xi r} \sin\theta \cos\phi$$

$$\chi_{3,0,0}^{STO}(\vec{r}) = \frac{2}{3\sqrt{10}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r}$$

$$\chi_{3,1,-1}^{STO}(\vec{r}) = \frac{2}{\sqrt{30}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \sin\theta \sin\phi$$

$$\chi_{3,1,0}^{STO}(\vec{r}) = \frac{2}{\sqrt{30}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \cos\theta$$

$$\chi_{3,1,+1}^{STO}(\vec{r}) = \frac{2}{\sqrt{30}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \sin\theta \cos\phi$$

$$\chi_{3,2,-2}^{STO}(\vec{r}) = \frac{1}{\sqrt{6}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \sin^2\theta \sin(2\phi)$$

$$\chi_{3,2,-1}^{STO}(\vec{r}) = \frac{2}{\sqrt{6}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \cos\theta \sin\theta \sin\phi$$

$$\chi_{3,2,0}^{STO}(\vec{r}) = \frac{1}{3\sqrt{2}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} (3\cos^2\theta - 1)$$

$$\chi_{3,2,+1}^{STO}(\vec{r}) = \frac{2}{\sqrt{6}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \cos\theta \sin\theta \cos\phi$$

$$\chi_{3,2,+2}^{STO}(\vec{r}) = \frac{1}{\sqrt{6}}\sqrt{\frac{\xi^7}{\pi}}\, r^2 e^{-\xi r} \sin^2\theta \cos(2\phi)$$

$$\chi_{4,0,0}^{STO}(\vec{r}) = \frac{1}{3\sqrt{35}}\sqrt{\frac{\xi^9}{\pi}}\, r^3 e^{-\xi r}$$

2.3. Molecular properties.

Many QMC methods are very good and efficient, especially for small molecules. They can provide us with the ground state non-relativistic energies which are comparable to the best Configuration Interaction (CI) results. They can even provide us with various molecular properties like the harmonic, anharmonic and the vibration-rotation constants, all related to the energy derivatives due to the bond distance. We can also get the dipole moment which is also related to the energy derivatives due to the electrostatic field imposed on the molecule but is actually calculated in a different way in the present approach. The above mentioned properties are differential properties always connected to a small parameter of perturbation present in the Hamiltonian which corresponds to either an infinitesimal stretch of the bond or to an infinitesimal electric field. So far the results for properties were presented only for small molecules, but in this work we show that the same thing can be done for molecules with large nuclear charge. Our approach is based on optimization of the valence part of the wavefunction to get better properties.

There are several QMC approaches treating the derivatives of the molecular energy. One of them is the Differential QMC [45] which involves direct estimates of the energy difference between infinitesimally close systems and then computes the molecular property as a numerical derivative. This was applied to the dipole moment of the lithium hydride molecule. Another technique [46] uses the straightforward differentiation of the local energy estimate obtained from the DQMC with respect to the given parameter and subsequently estimates these expectation values by Monte Carlo method. This method was applied to the investigation of the energy curve of H_2 for different interatomic distances. The third method, **Infinitesimal Differential Diffusion Quantum Monte Carlo** [47,48] estimates the first and higher order derivatives of any operator (not only

the Hamiltonian), based on the "mixed" distribution $\phi_0\psi_T$ (used in DQMC) rather than variational based on $|\psi_T|^2$. This method is closely related to the Differential QMC, but the derivatives are done **analytically**.

In paper [47] the Hellman-Feynman theorem was utilized. This theorem says that having the **exact** wave functions ψ_n we can compute the derivatives of the corresponding eigenvalues E_n due to the parameter present in the Hamiltonian in the following way

$$\frac{\partial E_n}{\partial \lambda} = \int \psi^*_n \frac{\partial H}{\partial \lambda} \psi_n \, d\tau \qquad (1)$$

In another words we compute (or estimate) the expectation value of another operator, namely $\partial H/\partial\lambda$ using the same eigenfunction ψ_n. This is true only for the exact WFs and certain classes of approximate WFs. For QMC it translates into necessity to sample from the exact distribution. The use of the Hellman-Feynman theorem (HFT) did not prove to be very useful. There were large systematic and statistical errors of the first and the second energy derivatives present. In this approach instead of branching the walkers, their weights were cumulated [49].

In paper [48] a new approach was used to avoid the use of the HFT. The trial WF ψ_T was made λ dependent (λ is stretching parameter of the bond) by moving the molecular orbitals along with the atom (without changing the molecular orbital coefficients). Time step $\Delta\tau$ was also λ dependent. The previous technique of replacing branching with weights was used as well. However, rather than having the squared weights as before, the ordinary w_i were employed to achieve DQMC sampling which had the fixed-node error only. This can be called "mixed" sampling (using $\phi_0\psi_T$) and is capable of providing good values for energy derivatives. All the calculations were performed for one internuclear separation and the stretching was done only by utilizing the derivatives of different quantities, the molecule was not stretched "physically". However, these derivatives were done numerically.

The above mentioned method estimates the expected value of an operator A by using a grand mean over many iteration averages A_i obtained as follows

$$<A_i> = \frac{\sum_{j=1}^{NS} A(\mathbf{R}_{j,i})\, w_j}{\sum_{j=1}^{NS} w_j} \qquad (2)$$

The summations are performed through the whole ensemble of NS configurations where $\mathbf{R}_{j,i}$ is the j-configuration "position vector" in the i-th iteration. The weights are defined in the following way

$$w_j = \prod_{i=1}^{\infty} \exp(-\Delta\tau(E_L(\mathbf{R}_{j,i})-E_0)) \qquad (3)$$

However, incorporating the λ dependency into the weights must be performed. Since we change both the potential energy and the trial WF ψ_T, these changes will have an effect on weights. Finally we will need just the derivatives of these weights computed for the fixed simulation bond distance (corresponding to λ=0). Taking all that into account we can express different derivatives of weights and the special averaged combinations (cumulants) of the above mentioned quantities. Because of the singularities of E_L and the w_i-values near the nodes of ψ_T it was necessary to perform truncation of these quantities to be able to obtain a reasonable averaging. All necessary formulas as well as other details of the actual calculations are explicitly shown in ref.[48].

The method was used for calculation of the first energy derivative $\partial E_0/\partial\lambda$, harmonic and anharmonic constants (ω_e,$\omega_e x_e$) along with the vibration-rotation interaction constant (α_e) for LiH and LiD. All the quantities were obtained with good precision. The first energy derivative no longer exhibits the systematic statistical error which was caused previously by using the Hellman-Feynman theorem. As we have mentioned before, we must have another criterion along with the valence variance and that is the value of the first derivative of the ground state energy due to the bond distance. We can either monitor it during the simulation of the system or

calculate it later having the values of coefficients and other parameters of the WF that we want to "check".

For the calculation of molecular properties of CuH a method rather similar to the one described in [47,48] was used. The main difference is that instead of using DQMC approach with the weights given by the formula (3), variational approach without weighting ($w_j = 1$) was used. The reason for that is connected to the problems with large-Z systems. Since the fluctuations of the energy are so huge, the procedure becomes unstable. The other difference is that the trial WF doesn't depend on the infinitesimal stretching of the bond distance. The reason is that the valence orbitals in the CuH molecule WF are not so sensitive to the hydrogen shift as they were in LiH. Once more a large-Z plays an important role.

Therefore for the energy's first derivative we don't have any cumulants in the expression and we can incorporate it into our variational simulation program very easily. It will have the following form

$$\frac{\partial E_0}{\partial \lambda} = \int \psi_T \frac{\partial V}{\partial \lambda} \psi_T \, d\tau \qquad (4)$$

where the potential energy derivative is

$$\frac{\partial V}{\partial \lambda} = - \frac{Z_{Cu}}{\Lambda^2} + \sum_{i=1}^{N} \frac{z_i + \Lambda}{(x_i^2 + y_i^2 + (z_i + \Lambda)^2)^{3/2}} \qquad (5)$$

where $\mathbf{r}_i$ is the i-th electron position vector from the center of the frame system situated in the Cu nucleus.

3.1. Details of the calculations.

As we have already mentioned, the system of interest is the copper hydride molecule (CuH). The method outlined can be used even for other heavy atom hydrides and maybe for some diatomic or even polyatomic molecules. If we want to stay within the transition metals, the next molecule to treat would be silver hydride (AgH) or gold hydride (AuH) because both Ag and Au have similar electronic structures to the Cu atom.

For the purposes of our computer simulation, CuH is just a set of two nuclear charges (one with $Z_{Cu}=29$ and the other with $Z_H=1$) separated by the bond distance $\Lambda=2.764$ a.u. . We have these two nuclei fixed, and in our simulations we move just the electrons. We assume that this Born-Oppenheimer approximation is not going to have an effect on properties of interest. The nuclei are interacting electrostatically with each other and with the moving electrons and the electrons are interacting with each other. The electrons are moved, but this movement is only **mathematical** and has nothing to do with the actual physical movement of the electrons, which moreover cannot be described classically by assigning coordinates to the electrons.

The molecule is placed into the reference frame as the following figure shows

where (x_i,y_i,z_i) are the Cartesian coordinates of the i-th electron position

vector $\mathbf{r}_i$, with respect to the Cu nucleus.

In the previous chapter we have mentioned how important the wave function is in our calculations, and now we will go into more detail on the WF. We have 15 atomic orbitals (AO's) centered on Cu (optimized by Clementi and Roetti [44]) and two AO's (1s and 2p0) centered on H. However, we do not take the Clementi and Roetti atomic orbital coefficients. We just use their basis set. To be more clear, for the first part of our investigation, our molecular orbital coefficients $C_{i\nu}$ (in eq.(2.2.5.)) are equal to Kronecker delta $\delta_{i\nu}$ ($C_{i\nu}=\delta_{i\nu}$) for all i,ν except for i=15 (we want to introduce some corrections to AO's). MO ϕ_{15}, in our case, is carrying the valence contributions from AO's with the center on hydrogen ($1s_H$ and $2p0_H$) to the valence (the most wide spread) atomic orbital with the center on copper ($4s_{Cu}$). One could immediately argue that the mixing of the 4s and 3d orbitals (mentioned in chapter 1.6) which is important for the transition metals is not present and therefore we can not pick up the properties correctly. The fact is that we want to replace rather complicated valence combinations with a very simple one which is, however, optimized. We have decided to consider the $4s_{Cu}$ AO to be of the highest importance for the valence. Our matrix of the MO coefficients for spin up and down electrons will look like this

	ATOMIC ORBITALS						
	$1s_{Cu}$	$2s_{Cu}$	...	$3d0_{Cu}$...	$4s_{Cu}$	$1s_H$	$2p0_H$
MO's							
1	1	0	...	0	0	0	0
2	0	1	...	0	0	0	0
3	0	0	...	0	0	0	0
.	0	0	...	0	0	0	0
10	0	0	...	1	0	0	0
.	0	0	...	0	0	0	0
14	0	0	...	0	0	0	0
15	0	0	...	0	1	****	****

where all the elements in Cu part except those on the diagonal are zeros. In the H part only elements with asterisks are nonzero. The same matrix is used for both spin up and spin down electrons which corresponds to what was said in chapter 2.2.

To illustrate the atomic orbitals used for valence as well as molecular orbital ϕ_{15} , we have depicted the **r**=(0,0,z) dependency on fig.3.1. and the **r**=(x,0,z) dependency on figures 3.2. and 3.3. (axes x, y, and z on the figure corresponds to z, x and the value of the orbital respectively).

We have previously mentioned the split tau approximation. This approximation will not have a significant effect on the results providing the electrons do not change their $\Delta\tau$ values too often. That means they stay within the same shell (with the same $\Delta\tau$ value) during the whole simulation. Now we present some numbers which show that the electrons don't change $\Delta\tau$ values too often. The table below gives the probabilities of "shell change" for individual electrons; e.g. the probability that the electron from the first shell will be assigned to the second shell in the next iteration is 0.0208.

from shell		**1**	**2**	**3**	**4**
to shell	1	0.9792	0.0052	0.0	0.0
	2	0.0208	0.9755	0.0085	0.0
	3	0.0	0.0193	0.9914	0.0011
	4	0.0	0.0	0.0001	0.9989

This shows that the electron changes its $\Delta\tau$ value only once in a while and this is not going to affect the quality of the simulation under Metropolis sampling.

As we have said before, we have to group the electrons into shells. Figure 3.4. shows what the average distances of different electrons are as well as their typical energies (values are in atomic units). These values are the averages computed using approximately 1000 configurations. One can clearly see the huge energy scale differences present in the system. The

Figure 3.1.
Atomic and Molecular Orbitals.

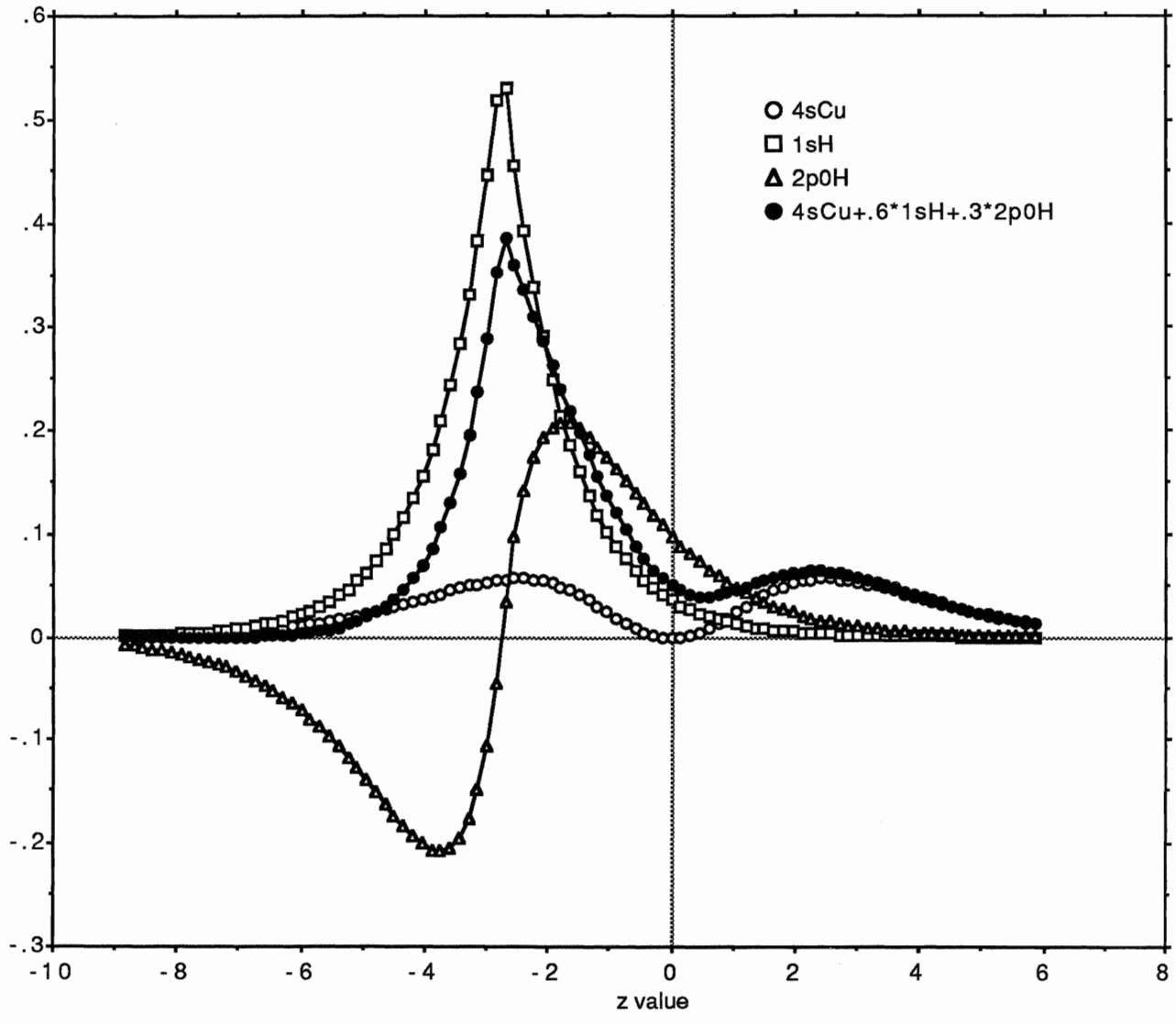


Figure 3.2. **Atomic Orbital $1S_H$ and $2p0_H$**

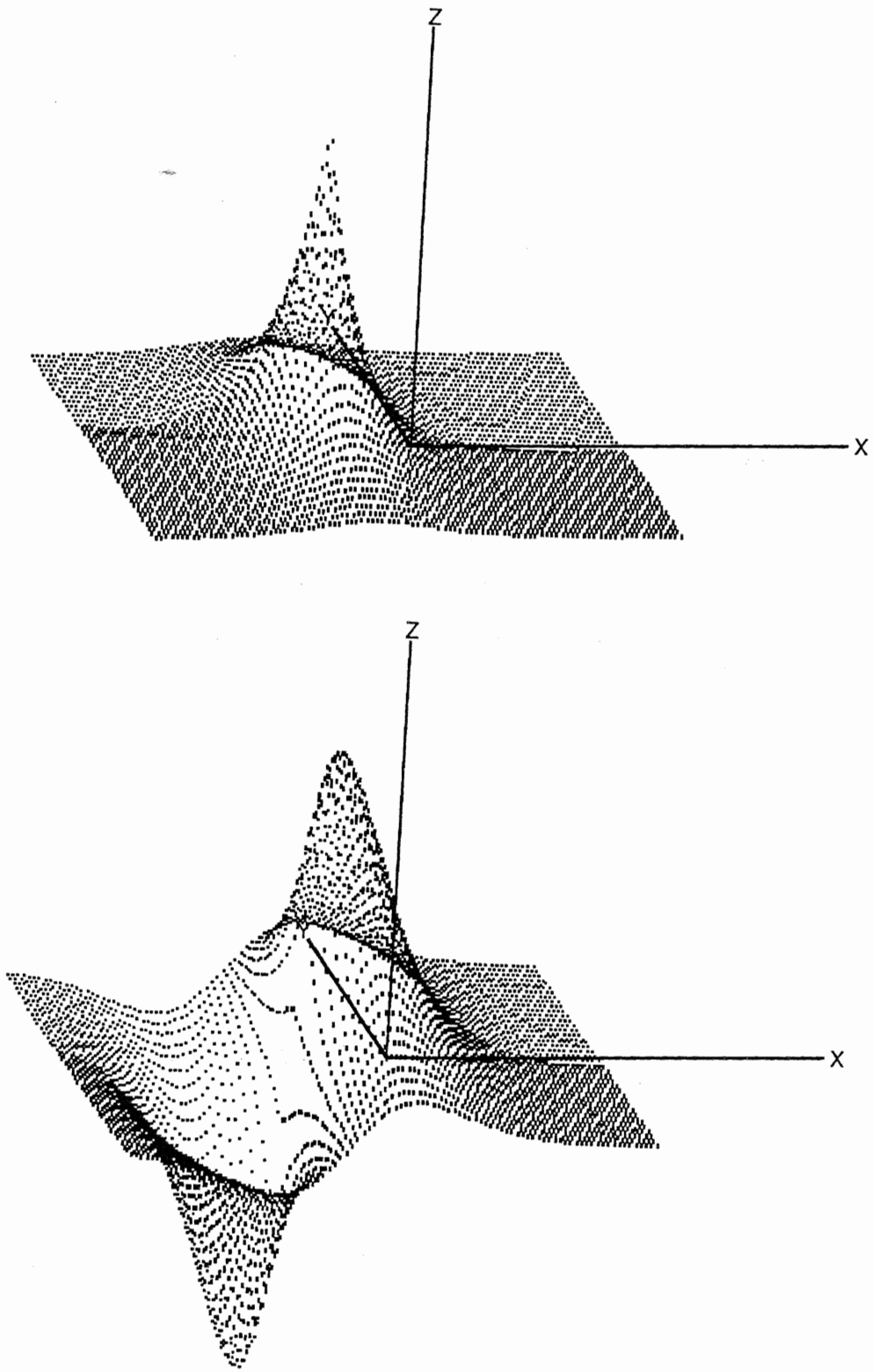
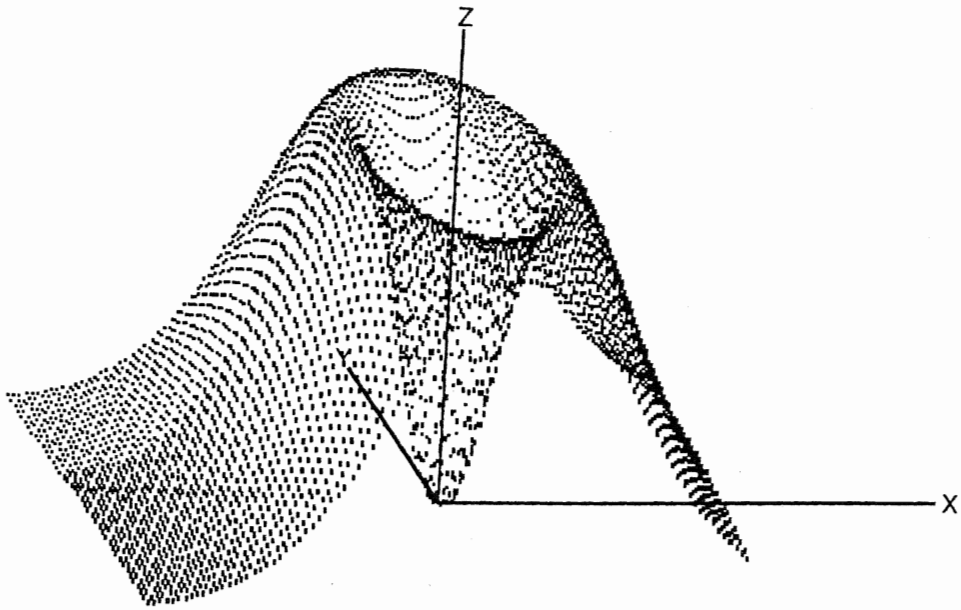
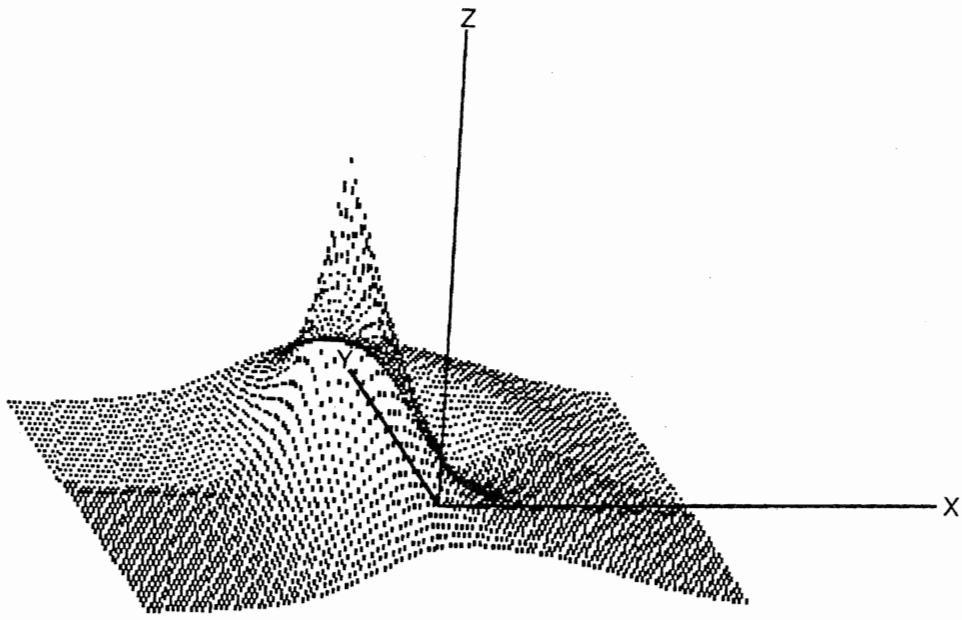


Figure 3.3. Atomic Orbital $4S_{Cu}$



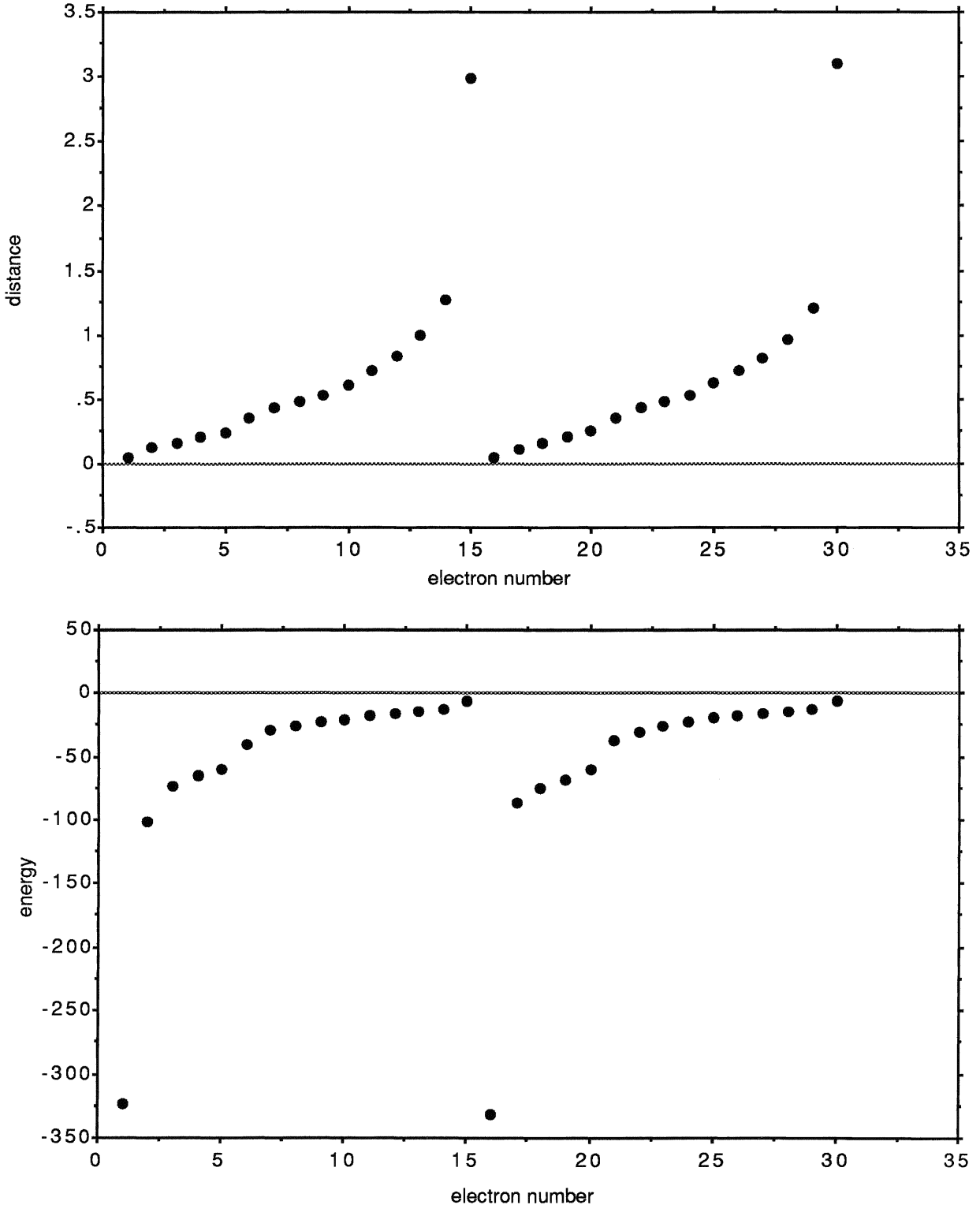
Molecular Orbital $4S_{Cu} + 0.6*1S_{H} + 0.3*2p0_{H}$



other thing which is worth noticing is that the two most distant spin up and spin down electrons (in our case denoted by labels 15 and 30) are clearly separated from the rest of the electrons. This immediately tells us that the outermost shell (valence) will consist of these two electrons only. One could think, that the rest of the shell split is arbitrary, because we take only valence electron energies into consideration and the rest is thrown out as an energy fluctuation background. This is, however, not quite true. The reason is obvious. Although the valence variance is not affected by the shell split of the rest of the electrons directly, it is affected by choice of $\Delta\tau$ values for them. The inappropriate time steps could introduce larger fluctuations into the valence and thus to the valence variance itself, preventing us from picking up the "valence parameters based" behaviour of valence variance, if any. There is a certain pattern between the distance of the electron, the molecular orbital "most appropriate" for the "description" of that particular electron, the $\Delta\tau$ value chosen and the shell we put it in. The values chosen are present in the following table

electron	MO	time step	shell
1(16)	1s	0.0001	1
2(17)		0.0005	2
3(18)	2s&2p	0.0005	2
4(19)	group	0.0005	2
5(20)		0.0005	2
6(21)		0.0012	3
7(22)	3s&3p	0.0012	3
8(23)	group	0.0012	3
9(24)		0.0012	3
10(25)		0.0040	3
11(26)	3d	0.0040	3
12(27)	group	0.0040	3
13(28)		0.0040	3
14(29)		0.0040	3
15(30)	4s	0.0040	4

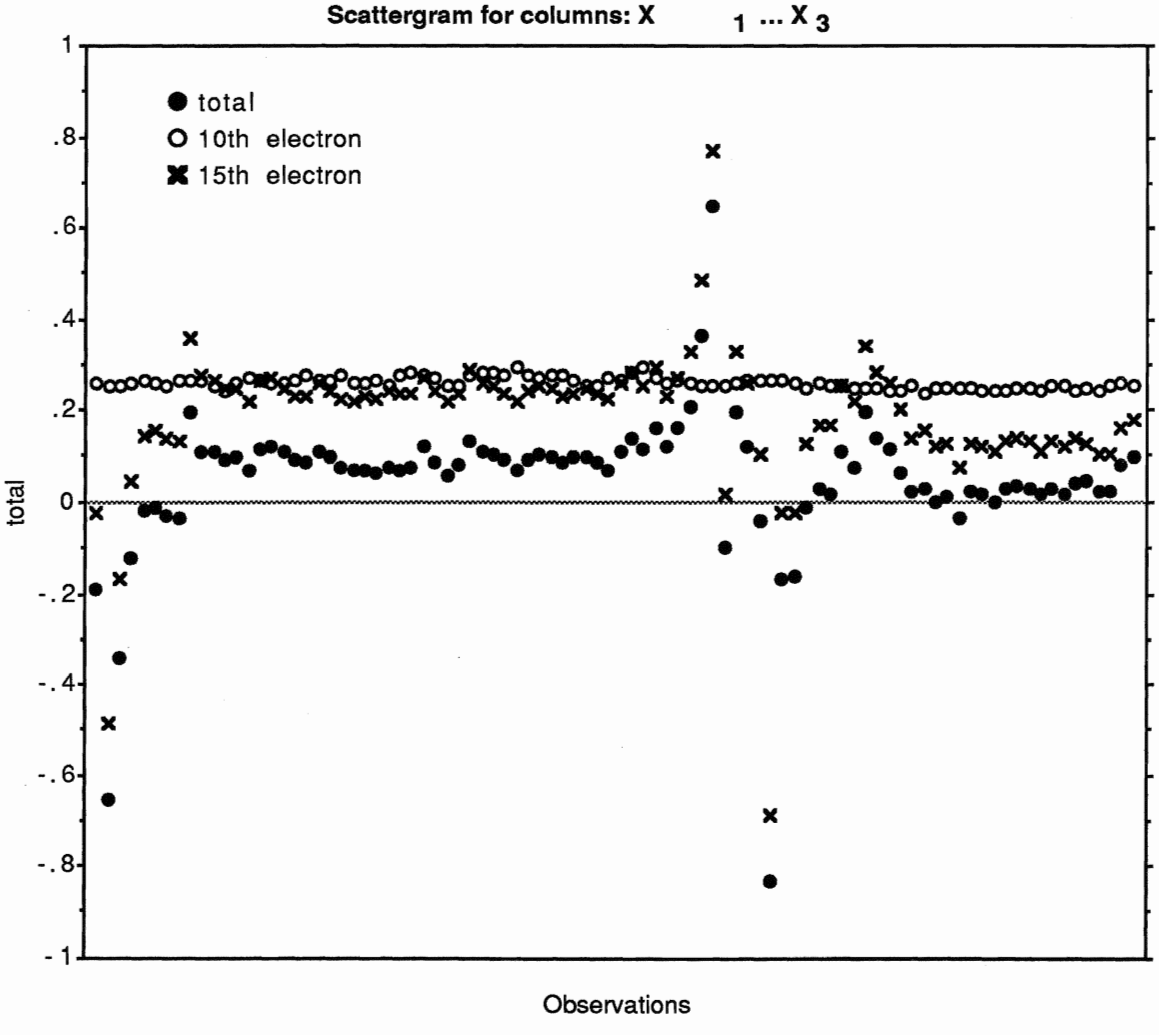
Figure 3.4.
Average Distance and Energy of the Electrons.



In connection with the split-tau, there is one other thing we have to mention. In each iteration step we assign $\Delta\tau$ values to the electrons according to their distances from the Cu nucleus, and the next move is done with the given $\Delta\tau$'s corresponding for each electron. This can be achieved without the necessity of ordering the electrons "physically" in the memory array. In spite of that, this procedure is done, because we want to make the inversion of the Slater determinant matrix more accurate. If we let the electrons correspond to the matrix columns in a random manner, once in a while we will have a very small value present on the diagonal of the matrix. Since we have to divide by these diagonal elements in our inverse subroutine, the procedure will become unstable and the matrix inverse will not be computed with desired precision. To avoid this, we order the electrons physically in ascending order of their distances from Cu nucleus. This ensures that there will be no extremely small elements on the diagonal, because the MO sequence we have for the Slater determinants (presented in the table 2.2.1.) is in ascending order of their average radials. Since we want as fast an inverse subroutine as possible, our subroutine doesn't have the pivoting feature implemented because it slows down the performance (the program spends a lot of CPU time doing the inverse). Instead of that we are doing the physical exchange of electrons which needs to be done only once in a while and therefore saves computer time.

As far as the formula for calculating the first derivative (eq.2.3.5.) is concerned, it is clearly seen that if we approach the Cu nucleus with the electron ($\mathbf{r}_i = \mathbf{0}$), then we have a $1/\Lambda^2$ contribution coming from this electron into the total first derivative. Fig.3.5. shows the contributions from different electrons into $\partial E_0/\partial\Lambda$. It is clear that the core electrons contributions are almost constant. The fluctuations are very small and the derivative value is always in the range 0.24 - 0.29 even for the tenth electron. This is, however, not true for the outermost electrons the values of which are fluctuating severely and therefore have a significant effect on the final value $\partial E_0/\partial\Lambda$.

Figure 3.5.
Contributions to the First Derivative from Different Electrons.



3.2. Valence variance for set of $\underline{s}$ and $\underline{p}$ values.

In this section we will present the results obtained for the valence variance. As well, we will suggest the neighbourhood of the parameters that should be investigated further. In our approach there are only two parameters : coefficients $C_{4s,1sH}$ and $C_{4s,2p0H}$ referred to as $\underline{s}$ and $\underline{p}$ in the text below.

The values of the variance will have large standard errors and the optimum $\underline{s}$ and $\underline{p}$ parameters will be probably very difficult to determine. Therefore, rather than using the optimization procedure described in chapter 1.4. by estimating the derivatives and making "steepest descent" steps in the parameter space (according to these derivatives), we decided to sample the $\sigma^2_{valence}$ value directly using quite long runs for each $(\underline{s},\underline{p})$ set of parameters. These results are presented in figures 3.6. to 3.9. What we can immediately see from these graphs is that the differences in some regions are indeed very difficult to notice.

In this case we don't have an explicit expression for the function we want to minimize (here $\sigma^2_{valence}(\underline{s},\underline{p})$) and the sampled means of valence variance have extremely high standard errors as well. What is usually done in the case like this is that we use multiple regression. We assume the simple behaviour of the minimized function $\sigma^2_{valence}(\underline{s},\underline{p})$ to be quadratic both in $\underline{s}$ and $\underline{p}$ parameters. We have tried to do the weighted as well as non-weighted regression on these data, but since the data are obviously not showing the desired dependency on $\underline{s}$ and $\underline{p}$ parameters, we were not able to find the minimum. The reason was that the coefficients appearing in our regression had very large standard errors. Taking into account these errors, we could get almost any combination of $\underline{s}$ and $\underline{p}$ to be the one which minimizes $\sigma^2_{valence}$. This is obviously not how it should be.

Since the multiple regression method was not working well in this case we had to use a less "sophisticated" method. We have to estimate the

minimum roughly by looking at the given graphs of valence variance. By doing so we can find two combinations of s and p showing the minimum of $\sigma^2_{valence}$. The combinations of s and p are (0.6,0.3) and (0.4,0.3) . In the next chapter we show the results for the properties obtained for these wave functions as well as some others.

In the graphs below we present the results of our simulation of the valence variance. Each point in the graphs is a mean taken from 50 block values of valence variance and the error bars are given as a standard errors for those means. Each block consists of 30 iterations and each iteration is calculated for the ensemble size of 30 configurations. To have an idea of how much CPU time is needed to sample the valence variance for these results we used about 6h of Silicon Graphics workstation CPU time for each single point. The computer used was an SGI Indigo workstation with the 33MHz RS3000 processor. Talking about CPU time, we must say something about the CPU requirements of the properties programs for different values of τ . The largest τ=5. value requires about 6.8h to run 20 blocks of 68 iterations (for the ensemble size of 100 configurations), whereas the smallest τ=1. for 20 blocks of 500 iterations requires approximately 50h of CPU time.

Since we want to sample the valence regions of the molecule in the best way, not paying too much attention to the core, there are ways to make this simulation faster. We can, for instance, **move only the two** valence electrons while leaving the rest at the same place for a while and only after we perform about 10-100 of these "valence" moves, we **move all** of the electrons again. This procedure will make it 5-20 times faster than in the case of moving all the electrons simultaneously. Unfortunately this brings another bias into the method and that is why we didn't do it. We wanted to avoid as much as possible bringing new unknown influences into the method already biased by using **split-τ** approximation. Nevertheless this might be a good field for further study.

Figure 3.6. Valence Variance for S=0.4 and different P.

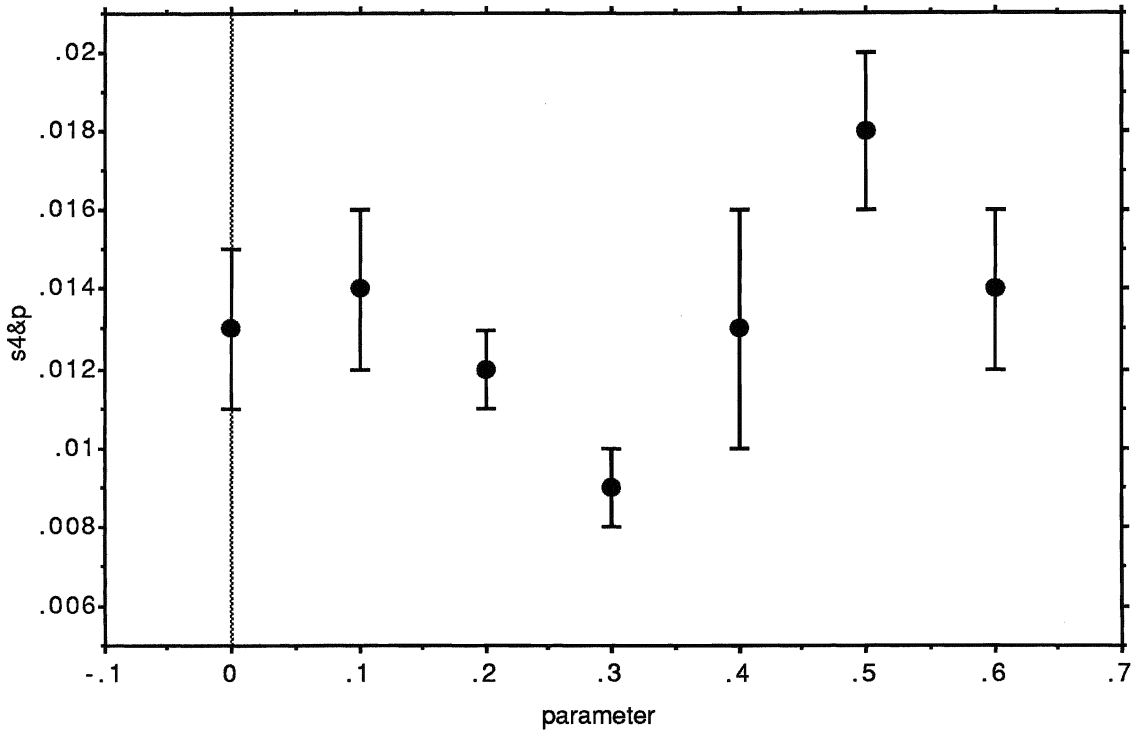


Figure 3.7. Valence Variance for S=0.5 and different P.

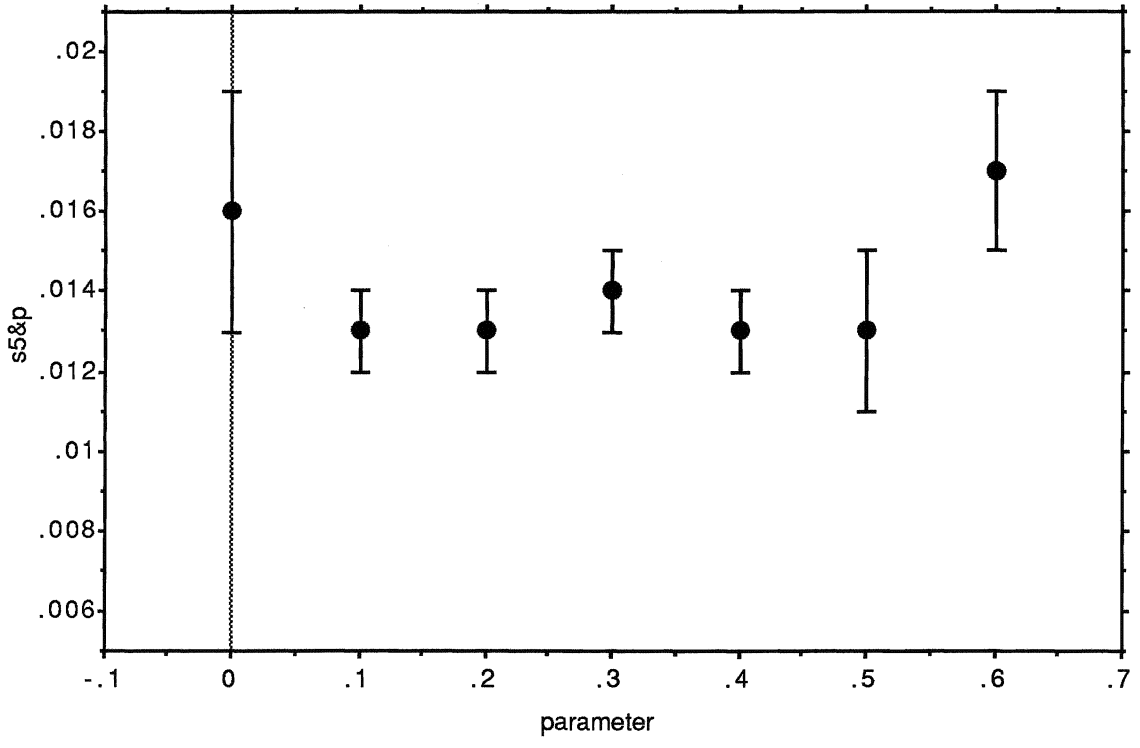


Figure 3.8. Valence Variance for S=0.6 and different P.

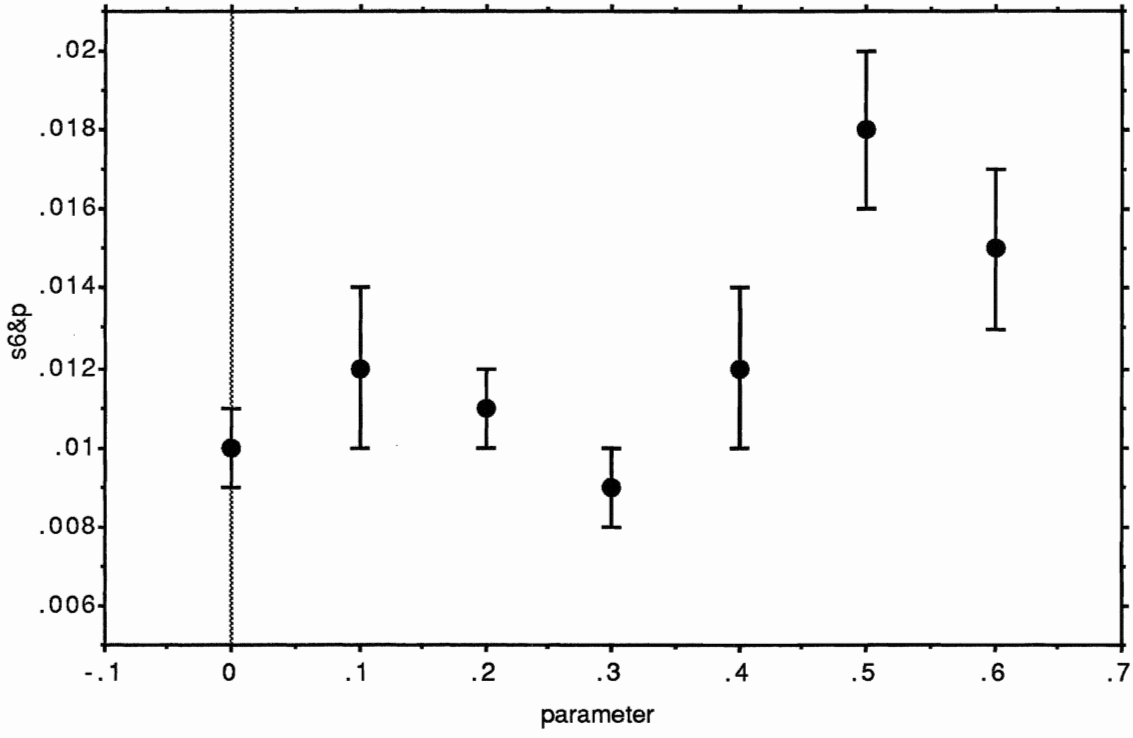
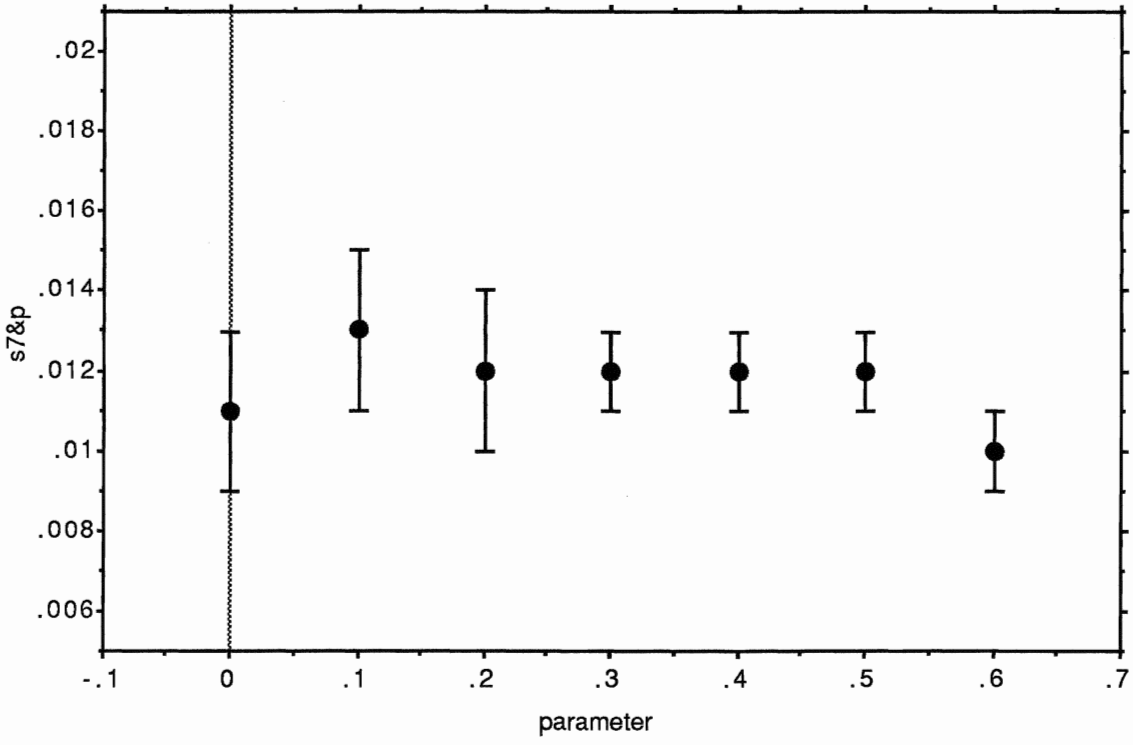


Figure 3.9. Valence Variance for S=0.7 and different P.



3.3. Properties for given wave functions.

In this chapter we will present the molecular properties calculated for different values of parameters (different wave functions). We will briefly describe why we have chosen such values of parameters to investigate, then we will describe the very important process of making the limit $\Delta\tau \rightarrow 0$ by doing regression. After this we present the molecular properties' values along with their uncertainties due to the statistical fluctuations for the given method of extrapolation.

As we have noted in the previous chapter, the parameters which are exhibiting lower valence variance than the rest are as follows:

s=0.6 , p=0.3

s=0.4 , p=0.3

In addition to these (as a sensitivity study), we have picked some other parameters like

s=0.6 , p=0.2

s=0.49 , p=0.24 and

s=0.5 , p=0.3

which are close to the first mentioned parameters and their valence variances are very close to the values of those "best variance" parameters as well.

The other parameter combination we have chosen is

s=0.7 , p=0.6

This choice has comparable valence variance, but it is very far from the other two groups of parameters we have investigated. In this case we wanted to see what the properties will be like for such a distant point in the parameter space.

The computer program for calculating the molecular properties, designed by Prof. J.Vrbik, follows the ideas mentioned in chapter 2.3. (for more details see [49]). The program calculates the first four derivatives of

the ground state energy of the molecule with respect to the bond distance and derives the spectroscopic constant from them. It computes the harmonic (ω_e) as well as anharmonic vibrational frequencies ($\omega_e x_e$) using the following formulas [50] :

$$\omega_e = \frac{1}{2\pi}\sqrt{\frac{E_0^{(2)}}{\mu_M}} \tag{1}$$

and

$$\omega_e x_e = \frac{1}{24}\left[\,5\,\left(\frac{E_0^{(3)}}{E_0^{(2)}}\right)^2 - 3\,\left(\frac{E_0^{(4)}}{E_0^{(2)}}\right)\right]\frac{1}{4\pi\mu_M} \tag{2}$$

where μ_M is the reduced mass of the system

$$\mu_M = \frac{M_{Cu}\,M_H}{M_{Cu} + M_H} \tag{3}$$

and the superscript on $E_0^{(i)}$ denotes the order of the derivative.

In addition to these two properties the distortion or vibration-rotation constant (α_e) [50] as well as dipole moment (μ the z component of the vector μ) are calculated using the following formulas

$$\alpha_e = -\left(\frac{E_0^{(3)}}{3\,E_0^{(2)}}R_e + 1\right)\frac{6\,B_e^2}{\omega_e} \tag{4}$$

where B_e is the rotational constant at the equilibrium intermolecular distance R_e

$$B_e = \frac{1}{2\mu_M R_e^2} \tag{4a}$$

and

$$\mu = \sum_i q_i\,r_i \tag{5}$$

The last molecular property which we report in this work and which is also given as a combination of the calculated derivatives is the equilibrium molecular distance (R_e). Throughout the calculations we used the experimental bond distance Λ =2.764 a.u.. For the exact wave function we should get the first derivative equal to zero. Since our wave function is only an approximation, we never get exactly zero for the first derivative. This allows us to make the correction of the bond distance based on $E_0^{(1)}$ and $E_0^{(2)}$ using **Newton-Raphson method**. We calculate the new equilibrium bond distance predicted by our MC method using the following formula

$$R_e = \Lambda - \frac{\frac{\partial E_0}{\partial \Lambda}}{\frac{\partial^2 E_0}{\partial \Lambda^2}} \qquad (6)$$

The method which is used here for the computation of molecular properties has the τ -bias. This is, however, going to disappear in the limit $\Delta\tau \rightarrow 0$ according to theory. For each wave function (set of parameters) all the properties were calculated (sampled) for five different τ values. Since τ are on a linear scale, we can denote them as τ =1.,2.,3.,4. and 5. This is connected to the split $\Delta\tau$ approximation we use so that the $\Delta\tau$ values used in different runs corresponding to the same shell are in the same proportion to each other as the τ values themselves. To have an idea of this, for the smallest τ run (τ=1.) the $\Delta\tau$ values were from 0.00016 (for the core electrons) to 0.01 (for the valence electrons) and for the largest τ run (τ=5.) the $\Delta\tau$ values were correspondingly from 0.0008 to 0.05 a.u.

We used 60 or 40 blocks per each τ run (slightly higher precision, using longer runs of 60, was achieved for those parameters showing better properties). However, the number of iterations was not the same for different τ values. The empirical formula [48]

$$N_{iter} = \left(3 + \frac{0.02}{\Delta\tau_{val}} \right) \frac{1}{\Delta\tau_{val}} \qquad (7)$$

was used for estimating the necessary number of iterations per block. The $\Delta\tau_{val}$ is the largest time step (valence $\Delta\tau$) for the given run. The number of iterations is increasing with decreasing τ value from 68 (for τ=5.) through 87, 122, 200 up to 500 (for τ=1.).

The most important thing in eliminating the τ-bias (τ=0. limit), is the proper estimate of the property value and its standard error. This would be no problem if we knew the functional dependence of the molecular property on τ (e.g. ω_e=f(τ)), but we don't. We have to proceed very carefully in the following manner. By checking the scattergram of the given property, we have to decide which model to choose so that we are neither overfitting the data nor approximating incorrectly. We had basically three possibilities which included the constant model (taking simply the weighted mean of all the data regardless of τ values when the data didn't exhibit any particular τ-dependence) and two polynomial models (simple linear and quadratic). The cubic model was already overfitting the data in the case of properties where it could be used in principle and it was not a good model for the other properties.

Now we will present the graphs of different properties along with the estimated intercept ($\tau \rightarrow 0$ limit value) and its standard error. In this graphical form we present only the properties obtained using s=0.4 and p=0.3 parameters whereas others are given in the tabular form for comparison. The solid points for different τ values on the graphs are the means of all the block values (60) and the error bars are given by standard errors of those means. Graph units for a particular property correspond to the table units. For some properties like ω_e and especially the dipole moment (μ) where large serial correlation was present, we accounted for this correlation in an usual manner by multiplying the standard error with the factor

$$\sqrt{\frac{1 + \rho(1)}{1 - \rho(1)}} \qquad (8)$$

where $\rho^{(1)}$ is the first order correlation coefficient.

On fig.3.11. $E^{(1)}$, $E^{(0)}$, ω_e , $\omega_e x_e$, α_e and μ graphs are depicted for the above mentioned wave function. For the ground state energy $E^{(0)}$ we report the offset from the -1640 a.u. value. As we have already mentioned, we used different regression models. For some of the properties like $E^{(1)}$, where it was not obvious which model to use, we have tried two or even all three of them and then enlarged the reported standard error for the intercept so that it accounted for the model bias as well. The other class of properties included $E^{(2)}$, ω_e and μ . These properties were giving us different τ point values that were spread out in a random manner and didn't show any visible functional dependency on τ . We have tried weighted averaging for them, but obviously this provided artificially low standard errors for the intercepts. We have decided to report the linear regression intercepts for them as well. However, the second derivative is not depicted graphically because it is very close to ω_e due to eq.(1). For the rest of the properties ($E^{(1)}$,$E^{(0)}$,$\omega_e x_e$,α_e) we report the quadratic regression intercepts calculated using a simple program doing a weighted quadratic regression.

When we are using the weighted regression for the set of values with significantly different standard errors, we usually end up not taking into account the values with largest error bars. This was the case of α_e as well as ω_e where the τ=1. point was nearly ignored.

One thing should be mentioned about the anharmonic constant $\omega_e x_e$. As we can see from eq.(2), this property is derived using third and fourth derivatives of the energy due to the bond distance . Since the method used for the estimating these derivatives is only an approximation, it has its own limitations. In the case of $\omega_e x_e$ we know that method exhibits the singularity for this property as $\tau \rightarrow 0$. We can "exclude" the smallest τ from our regression by doing the weighted regression and we get closer estimates for that property than doing non-weighted regression. Since this is not the proper way how to overcome the singularities which are coming out of the theory, the only thing we can conclude is that properties going beyond the third (or even the second) derivative have so far only the qualitative meaning, but are not quantitative.

In tables 3.4.a-f ω_e , $\omega_e x_e$ and α_e are expressed in cm^{-1} and all the other values are in atomic units (a.u.). In these tables the intercepts are the estimated values at $\tau=0$. The regression introduces a "model bias". There are statistical errors originating in the method as well as the "model bias". In the tables below we increased the statistical error, so that it accounts for the model bias as well.

Table 3.4.a. (s=0.6,p=0.3)

	intercept	st.error.
$E^{(1)}$	+0.011	0.007
$E^{(2)}$	0.173	0.003
E+1640	2.4	0.2
ω_e	2125.	19.
$\omega_e x_e$	44.	2.
α_e	0.12	0.01
dip.moment	1.10	0.05
R_e	2.70	0.04

Table 3.4.b. (s=0.6,p=0.2)

	intercept	st.error.
$E^{(1)}$	-0.028	0.009
$E^{(2)}$	0.192	0.005
E + 1640	2.4	0.2
ω_e	2245.	33.
$\omega_e x_e$	37.	2.
α_e	0.06	0.01
dip.moment	1.06	0.08
R_e	2.91	0.05

Table 3.4.c. (s=0.4,p=0.3)

	intercept	st.error.
$E^{(1)}$	-0.001	0.006
$E^{(2)}$	0.125	0.003
E + 1640	2.7	0.2
ω_e	1814.	25.
$\omega_e x_e$	50.	2.
α_e	0.23	0.01
dip.moment	1.59	0.04
R_e	2.77	0.05

Table 3.4.d. (s=0.7,p=0.6)

	intercept	st.error.
$E^{(1)}$	+0.048	0.004
$E^{(2)}$	0.156	0.003
E + 1640	2.9	0.2
ω_e	2021.	22.
$\omega_e x_e$	105.	6.
α_e	0.43	0.02
dip.moment	1.10	0.09
R_e	2.46	0.03

Table 3.4.e. (s=0.49,p=0.24)

	intercept	st.error.
$E^{(1)}$	-0.015	0.003
$E^{(2)}$	0.159	0.005
E + 1640	2.4	0.2
ω_e	2042.	32.
$\omega_e x_e$	38.	3.
α_e	0.09	0.02
dip.moment	1.13	0.06
R_e	2.86	0.02

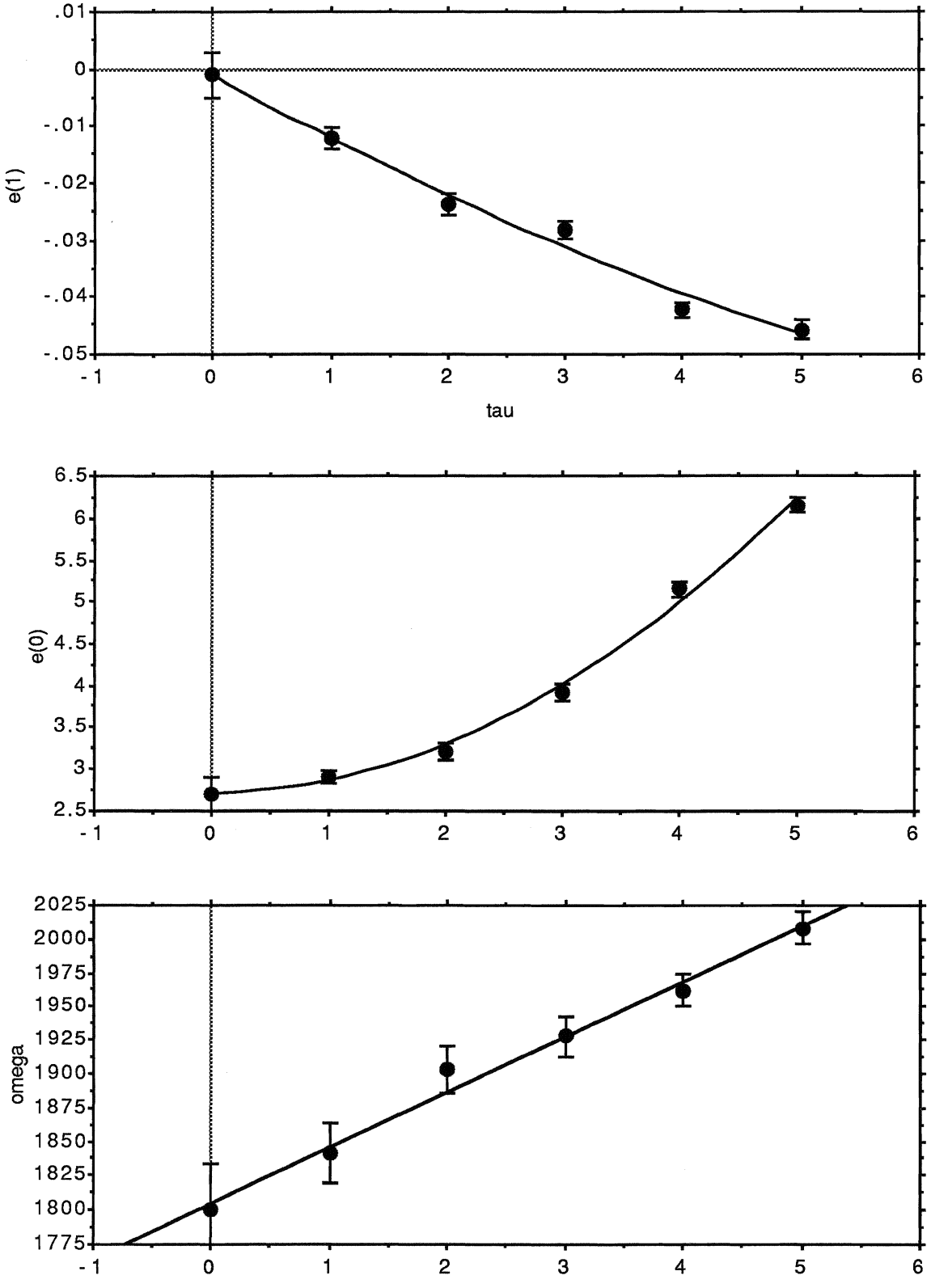
Table 3.4.f. (s=0.5,p=0.3)

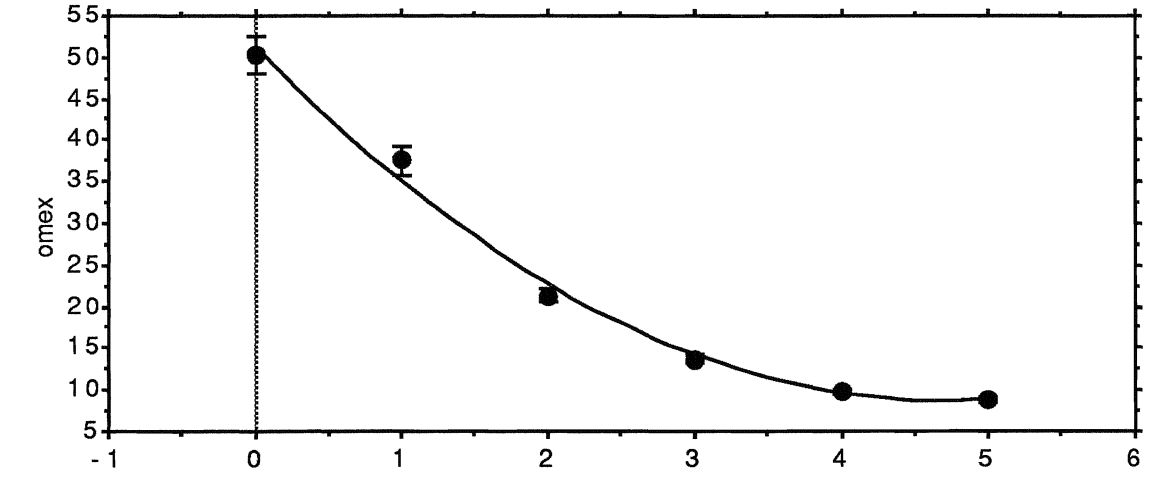
	intercept	st.error.
$E^{(1)}$	+0.014	0.008
$E^{(2)}$	0.161	0.006
E + 1640	2.3	0.4
ω_e	2058.	37.
$\omega_e x_e$	44.	4.
α_e	0.19	0.04
dip.moment	1.06	0.07
R_e	2.68	0.05

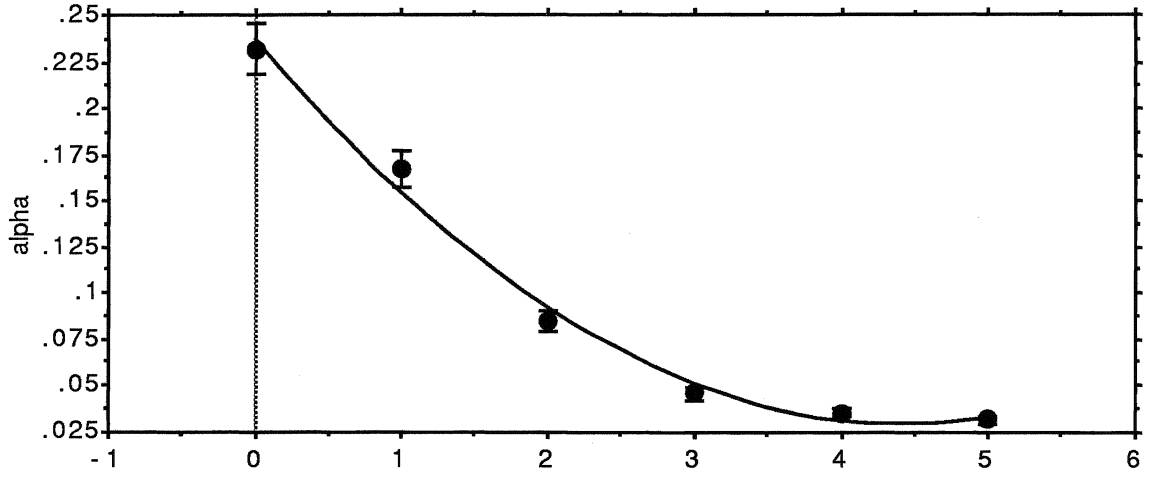
We know the experimental results for the properties of interest. These are given in K.P.Huber and G.Herzberg [Ref.51]. It might be surprising, but we didn't find the experimental value for the dipole moment in the literature. Therefore we had to take the best theoretical result given by C.M.Marian in [Ref.41]. He reports μ =1.2 a.u. for the internuclear separation equal to the experimental value. He used multireference singles and doubles configuration interaction (MRSDCI). The best available values for the properties are given in the table below.

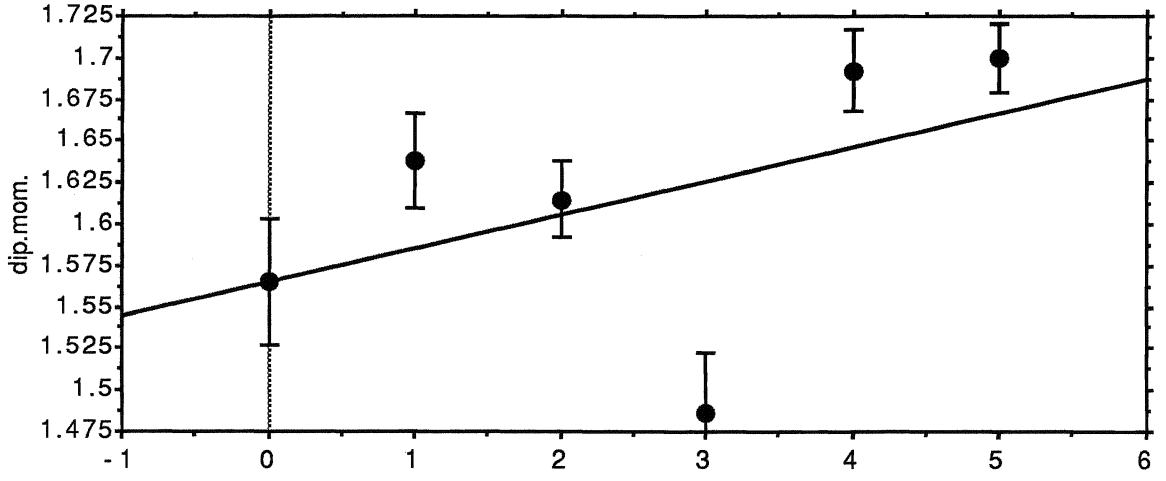
ω_e	= 1941. cm^{-1}	[ref.51] experimental
$\omega_e x_e$	= 37.5 cm^{-1}	[ref.51] experimental
α_e	= 0.2563 cm^{-1}	[ref.51] experimental
μ	= 1.20 a.u.	[ref.40] best theoretical
Λ	= 2.764 a.u.	[ref.51] experimental

Figure 3.11. Molecular Properties for S=0.4 and P=0.3




omex
55
50
45
40
35
30
25
20
15
10
5
-1
0
1
2
3
4
5
6


alpha
.25
.225
.2
.175
.15
.125
.1
.075
.05
.025
-1
0
1
2
3
4
5
6


dip.mom.
1.725
1.7
1.675
1.65
1.625
1.6
1.575
1.55
1.525
1.5
1.475
-1
0
1
2
3
4
5
6

4. Discussion.

In this work I have investigated the possibility of optimization of the copper hydride wave function based on the valence variance. We wanted to find the parameters $\underline{s}$ and $\underline{p}$ which minimize $\sigma^2_{valence}$ and we believed that this point (or possibly more points) will give us better properties than some other points randomly chosen from the parameter space. The reason we believed this approach could work is that the true WF should be somehow stable as opposed to the WF which "does not fit" the molecule. Using the more appropriate valence part for the description of the valence electrons must be reflected in the stability of the valence energy. The quantity which reflects this stability is the valence variance and having an inappropriate WF means more fluctuations of the energies of the simulated electrons. This holds exactly for the fully antisymmetric WF describing all the electrons. In that case having the **exact** WF gives us the exact value for the energy (in our case E_L) for **any** configuration of electrons. If there was no correlation between electrons, we could solve the problem very easily by looking at the each electron separately or in other words by separation of variables. As was mentioned earlier, we can not, however, isolate those two valence electrons from the rest, but we can expect that since the outermost electrons are so distant from the rest, the fluctuations of the energy will smooth each other out and we will be able to see whether the valence energy has stabilized or not. Therefore having a sufficient amount of computer time would allow us to find the real behaviour of $\sigma^2_{valence}$ as a function of parameters.

Following this and using the "inspection" method (because the regressions were not working well due to the large statistical error), we found two points that were significantly lower in valence variance value than the rest. We have calculated molecular properties for these two points

(s=0.6,p=0.3) and (s=0.4,p=0.3). The results are presented in the tables 3.4.a and 3.3.c. Both have the first derivative closer to zero than the rest, which results in very good values for the calculated equilibrium bond distances. However, only for (0.4,0.3) the exact value (namely zero) falls within the statistical error of our calculated value. For both points we have much better R_e than those achieved by other techniques like SDCI (2.853 a.u.), CPF (2.852 a.u.) or MCPF (2.851 a.u.) [40]. This is not true for the very close point (0.6,0.2) which gives us 2.91 a.u. as well as (0.7,0.6) with 2.46 a.u. Not using the exact bond distance for the calculations would probably give us a slightly worse value for R_e. The other methods (like SDCI) use several points of different bond distance to calculate the energy, then approximate the points by curve and calculate the equilibrium point from all these data. We can only use one geometry because of the large CPU time requirements for the calculations.

As far as the other properties are concerned, we think that both the harmonic frequency ω_e as well as the dipole moment μ calculated by the QMC method (this work) are comparable to those obtained by singles and doubles configuration interaction (SDCI) [Chong et al.,Ref.40]. The dipole moment is comparable within the statistical accuracy of our method as well. Unfortunately, the value for anharmonic constant $\omega_e x_e$ and α_e are not given for SDCI and therefore we cannot compare these. As we go to the much more sophisticated method MRDCI, we have to admit that QMC is not capable (at this stage) to give us values of comparable quality for the properties. As was pointed out before, $\omega_e x_e$ is biased because of singularities which are present for small τ and to obtain more precise results, these singularities have to be taken care of more properly. Below we present the table of the properties of interest given by the mentioned methods as well as the available experimental values.

		QMC	SDCI	MRD-CI	expt.
R_e	[a.u.]	2.77(5)	2.85	2.77	2.764
ω_e	[cm^{-1}]	1814(25)	1815	1952	1941
$\omega_e x_e$	[cm^{-1}]	50(2)	~	35	37.5
α_e	[cm^{-1}]	0.23(1)	~	~	0.2563
μ	[a.u.]	1.59(4)	1.53	1.2	~

From the two possibile parameter combinations we obtained by optimization, we have chosen the one which gives (or better confirms) the experimental geometry. The only problem is the dipole moment. Though we have almost the same value given by SDCI, it is far from the best CI result.

The work reported here is at the state-of-the-art for large-Z optimization. If one realizes, however, how crude is the wave function we used, building up the core part only using the atomic orbitals without mixing them together by introducing molecular orbital coefficients, as well as creating the valence part of the WF using only two adjustable parameters, then this optimization must be deemed a "success". We would be able to get better properties by having larger flexibility of the valence WF if we introduce more orbitals. We could probably improve the WF somewhat by optimizing the core part as well. We believe that the molecular environment by which the atom is surrounded has only a very small effect on the deepest shells of the atoms, therefore we could have the core part of the WF given once and for all and use it for that particular atom in any other molecule. The only changes which have to be implemented would take place only in the valence part of the WF, and we could do it by the optimization presented in this work (or something similar to this one). This is the similar philosophy as that of core potential with the difference of explicitly taking the Fermi correlation into consideration.

There is yet another advantage of our approach over the pseudopotential approaches. In QMC we are not approximating the core electrons by some potentials thus averaging their effect on the rest of the molecule. We simulate them all, though different electrons using different $\Delta\tau$ values, and therefore we are taking into account the correlation between core and valence.

As we have pointed out a couple of times in the previous chapters, we are dealing with the large-Z problem. For solving that, we have introduced the so called split-τ approximation dealing with all the electrons in the same variational way, which is naturally emerging from the fact that the electrons which belong to different shells have different sizes of their configuration spaces and therefore to sample them requires different τ values. We do the sampling using Variational Quantum Monte Carlo. The only potential bias

introduced into the sampling is due to the mentioned split-τ approximation and is connected to the electrons changing their shells. We have shown that this is not a very severe problem, because the electron shell interchange does not occur too often. The bias can be removed by going to the $\Delta\tau \rightarrow 0$ limit.

This work was the first attempt (of which we are aware) to calculate the properties of such a large molecule (30 electrons) by QMC using an all electron approach. This can be investigated further, but even this first step using a very simple WF is showing promise for such calculations in the future.

5. References:

[1] Symposium on Monte Carlo Methods, ed. by H.A.Meyer (Wiley - Interscience, New York , 1956)

[2] N.Metropolis and S.M.Ulam : J.Am.Stat.Assoc.**44**, 247 (1949)

[3] M.H.Kalos : Phys.Rev.**128**, 1891(1962)

[4] M.H.Kalos : J.Comput.Phys.**1**, 257(1966)

[5] J.B.Anderson : J.Chem.Phys.**63**, 1499(1975); J.Chem.Phys.**65**, 411(1976)

[6] F.Mentch and J.B.Anderson : J.Chem.Phys.**74**, 6307(1981)

[7] J.W.Moskowitz et al. : J.Chem.Phys.**77**, 349(1982)

[8] D.M.Ceperley and M.H.Kalos in : *Monte Carlo Methods in Statistical Physics*, ed. by K.Binder (Springer Verlag, Berlin, 1979), Chap.4

[9] B.J.Rosenberg, I.Shavitt : J.Chem.Phys.**63**, 2162 (1975)

[10] E.R.Davidson in : *Reviews in Computational Chemistry* (vol.1.), ed. by K.B.Lipkowitz and D.B.Boyd (VCH Publishers Inc., New York, 1990), Chap.11, p.373

[11] L.Hammond, P.J.Reynolds and W.A.Lester Jr. : J.Chem.Phys.**87**, 1130(1987)

[12] D.M.Ceperley : J.Stat.Phys.**43**, 815(1986)

[13] H.Gould and J.Tobochnik : Computers in Physics (jul/aug 1989)

[14] M.H.Kalos, D.Levesque and L.Verlet : Phys.Rev.**A9**,2178(1974)

[15] P.J.Reynolds, D.M.Ceperley, B.J.Alder and W.A.Lester Jr. : J.Chem.Phys.**77**, 5593(1982)

[16] R.J.Harrison and N.C.Handy : Chem.Phys.Lett.**113**, 257(1985)

[17] C.J.Umrigar, K.G.Wilson and J.W.Wilkins : Phys.Rev.Lett.**60**, 1719(1988)

[18] H.Bueckert, S.M.Rothstein and J.Vrbik : Can.J.Chem.**70**, 366(1992)

[19] R.N.Barnett, P.J.Reynolds and W.A.Lester Jr. : J.Chem.Phys.**82**, 2700(1985)

[20] P.J.Reynolds, R.N.Barnett, L.Hammond and W.A.Lester Jr. : J.Stat.Phys. **43**,1017(1986)
[21] F.Mentch and J.B.Anderson : J.Chem.Phys.**80**, 2675(1984)
[22] D.R.Garmer and J.B.Anderson : J.Chem.Phys.**86**, 4025(1987); J.Chem.Phys.**86**, 7237(1987)
[23] M.Hurley and P.A.Christiansen : J.Chem.Phys.**86**, 1069(1987)
[24] P.A.Christiansen : J.Chem.Phys.**88**, 4867(1988)
[25] T.Yoshida and K.Iguchi : J Chem.Phys.**88**, 1032(1988)
[26] H.Hellman : J.Chem.Phys.**3**, 61(1935); J.Chem.Phys.**4**, 325(1936)
[27] P.Gombas : Z.Phys.**94**, 473(1935)
[28] J.Carlson,J.W.Moskowitz and K.E.Schmidt : J.Chem.Phys.**90**, 1003(1989)
[29] J.C.Phillips and L.Kleinman : Phys.Rev.**116**, 287(1959)
[30] V.Bonifacic, S.Huzinaga : J.Chem.Phys.**60**, 2779(1974)
[31] S.Huzinaga, L.Seijo, Z.Barandiaran, M.Klobukowski :. J.Chem.Phys.**86**, 2132(1987)
[32] G.B.Bachelet, D.M.Ceperley and M.G.B.Chiocchetti : Phys.Rev.Lett.**62**, 2088(1989)
[33] B.L.Hammond, P.J.Reynolds and W.A.Lester Jr. : Phys.Rev.Lett.**61**, 2312(1988)
[34] D.P.Chong , S.R.Langhoff : J.Chem.Phys.**84**, 5606(1986)
[35] K.Raghavachari, G.W.Trucks : J.Chem.Phys.**91**, 1062(1989); J.Chem.Phys.**91**, 2457(1989)
[36] R.J.Gdanitz, R.Ahlrichs : Chem.Phys.Lett.**143**, 413(1988)
[37] C.W.Bauschlicher : J.Phys.Chem.**92**, 3020(1988)
[38] S.P.Walch, C.W.Bauschlicher, S.R.Langhoff : J.Chem.Phys.**83**, 5351(1985)
[39] C.M.Marian, M.R.A.Blomberg, P.E.M.Siegbahn : J.Chem.Phys.**91**, 3589(1989)
[40] D.P.Chong, S.R.Langhoff, C.W.Bauschlicher, S.P.Walch, and H.Partridge : J.Chem.Phys.**85**, 2850(1986)
[41] C.M.Marian : J.Chem.Phys.**94**, 5574(1991)
[42] R.Jastrow : Phys.Rev.**98**, 1497(1955)
[43] J.C.Slater : Phys.Rev.**36**, 57(1930)
[44] Clementi, Roetti : Atomic and Nuclear Data Tables **14**, 177(1974)

[45] B.H.Wells : Chem.Phys.Lett.**115**, 89(1985)
[46] P.J.Reynolds et al. : Int.J.Quant.Chem.**29**, 589(1986)
[47] J.Vrbik,D.A.Legare and S.M.Rothstein : J.Chem.Phys.**92**, 1221(1990)
[48] J.Vrbik and S.M.Rothstein : J.Chem.Phys.**96**, 2071(1992)
[49] A.L.L.East, S.M.Rothstein and J.Vrbik : J.Chem.Phys.**89**, 4880(1988)
[50] G.W.King : *Spectroscopy and Molecular structure*. (Holt, Rinehart and Winston, New York, 1964), Chap.5, p.161-164
[51] K.P.Huber and G.Herzberg : *Molecular spectra and molecular structure*, (Van Nostrand Reinhold,New York,1979)
[52] N.Metropolis, A.W.Rosenbluth, M.N. Rosenbluth, A.H.Teller and E.Teller : J.Chem.Phys.**21**, 1087(1953)
[53] C.W.Bauschlicher in : *Transition Metal Hydrides*, ed. by A.Dedieu (VCH Publishers, N.Y., 1991)